THE UNIVERSITY *of* EDINBURGH

Structural basis for iron (II) metabolism in encapsulated ferritin-like proteins

A thesis

Submitted for the Degree of

Doctor of Philosophy

by

Didi He, MSc (BSc)

The Institute of Quantitative Biology, Biochemistry and Biotechnology

School of Biological Sciences

The University of Edinburgh

Scotland

United Kingdom

Feb 2017

Abstract

Ferritins are ubiquitous proteins that serve the dual-function of iron reservoir and sequestering the Fe(II) toxicity. The function of ferritins totally depends on the characteristic spherical structure with a di-iron centre performing the iron oxidation and a hallow cavity enclosing the iron minerals in a bioavailable form. I have characterised the structure, assembly and function of a new member of ferritin superfamily that is natively enclosed within an encapsulin shell. Encapsulin proteins are structurally-related to a virus capsid and form 60-meric or 180-meric icosahedrons. I show that this encapsulin associated ferritin-like protein (EncFtn) possesses two main alpha helices, which assemble in a metal-dependent manner to form a ferroxidase centre at a dimer interface. EncFtn adopts an annular decamer structure in contrast to the 24-meric classical ferritins or 12-meric mini-ferritin (DPS). The resemblance of the dimeric EncFtn and monomeric classical ferritins suggests that it is likely that classical ferritin evolves from EncFtn because of the gene duplication. EncFtn is a catalytically active ferroxidase but with only a limited iron binding ability due to its open structure. The encapsulin itself is not able to oxidise Fe(II), but is able to store about 2200 iron ions. I have demonstrated that the EncFtn must be housed in the encapsulin to achieve a maximum loading of approximately 4200 iron ions. The encapsulin nanocompartments are widely-distributed in both eubacteria and archaeon with distinct life styles and represent a distinct class of iron storage system, where iron oxidation and mineralisation are distributed between two proteins.

Lay Summary

Iron as a basic element is essential for most living cells. But extra free iron in cells could damage cellular components. Nature has evolved a protein called ‘ferritin’ to maintain the iron concentration in a reasonable range in cells. Ferritins have a ball-like structure and enclose a central cavity where iron can be safely stored away from the rest of cells. I am interested in a new family of ferritins called ‘ferritin-like protein’, which has no clear information regarding its structure and function. My PhD thesis focused on three questions: 1. What is the protein structure of this ferritin-like protein? 2. What does ferritin-like protein do in cells? 3. Does the protein structure justify its function?

Declaration

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and has not been submitted in whole or in part for any other degree.

Didi He

Acknowledgements

I would like to express my sincere gratitude to my former first supervisor Dr. Jon-Marles Wright for his continuous support throughout my PhD training. Jon is the best supervisor and mentor that is willing to contribute his time, motivation, and immense knowledge to guide me on the correct track of my PhD research. He has helped me in almost all ways that I can think about. He has helped me improve my experiment designing skills with better organising ability and vigilant attitude towards data analysis. He has encouraged me to attend seminars or conferences and helped me improve my presentation skills. Even after Jon moved to Newcastle University in the end of my third year, he has never stopped guiding me through my writing and giving wonderful suggestions on my future career.

I would like to thank Dr. Janice Bramham who became my first supervisor after Jon moved back to Newcastle. She has been very supportive and offered me a demonstrating job to finance my thesis-writing year. I would like to thank my second supervisor Prof. Malcolm D. Walkinshaw for his precious encouragement and insightful suggestions. I also thank my thesis committee members Dr. Louise Horsfall, Dr. Julia Richardson and Dr. Patrick Cai.

My sincere thanks go to our collaborators Dr. David J. Clarke, David's students Sam Huges and Sally Vanden-Hehir from Chemistry School of the University of Edinburgh, Dr. Emma Tarrant and Dr. Kevin J. Waldron from Newcastle University. We have been in a close collaboration and exchanged fantastic research ideas. Special thanks go to Dr. Mike Capeness in Dr. Louise Horsfall's lab for the anaerobic glove box usage.

My sincere thanks also go to Atanas Georgiev. If it was not for supervising him, I probably would have never started this bacterial nanocompartment project. Thanks to Dr. Kirsten Altenbach who had contributed a lot to the lab management and provided fantastic molecular biology training. Great thanks to my labmates Laura Tuck and Efrain Zarazua-Arvizu for their great encouragement and the stimulating discussions. Laura has been very helpful on reading and revising my manuscript.

I would like to thank all people in Swann level 3 labs, especially scientists from EPPF including Dr. Liz Blackburn, Dr. Martin Wear and Dr. Matt Nowicki and Mrs. Sandra Bruce. I would like to thank Dr. Andrew Cronshaw for the mass spectrometry training. I would like to thank Dr. Steve Mitchell for all the guidance on transmission electron microscopy. I would like to express my gratitude to Prof. Bettina Böttcher for the cryo-EM workshop training. I would like to thank the Diamond Light Source facilities and CCP4 communities; without them I could not determine the protein structures. My deepest gratitude goes to the University of Edinburgh and my PhD scholarship sponsor—the Chinese Scholarship Council; without the funding I could not come to UK and finish my PhD study.

Thanks to Wellcome Trust-funded TGIF parties for providing great chances to do networking. Thanks to all my best friends who have looked after me physically and spiritually.

Last but not the least, I would like to thank my dear extended family, especially my mother and father.

Table of Contents

Abbreviation list

ACN	Acetonitrile
A_{280}	Absorbance at 280 nm
A_{315}	Absorbance at 315 nm
ABC	Ammonium bicarbonate
AEC	Anion exchange chromatography
AGF	Analytical gel-filtration
BFR	Bacterioferritin
BSA	Bovine Serum Albumin
CaAc	Calcium acetate
Chi2 angle of Trp	CA-CB-CG-CD1 of tryptophan residue
Chi3 angle of Glu	CB-CG-CD-OE1 of glutamic acid residue
DPS	DNA-binding protein from starved cells
DTT	Dithiothreitol
E. coli	*Escherichia coli*
Enc	Encapsulin
EncFtn	Encapsulin associated ferritin-like protein
FgMF	Bullfrog M-chain ferritin
FOC	Ferroxidase centre
FTN	Bacterial ferritin
H. ochraceum	*Haliangium ochraceum*
H. pylori	*Helicobacter pylori*
Hoch-EncFtnS	Hoch_$3836_{1\text{-}98}$

HuHF	Human H-chain ferritin
ICP-MS	Inductive coupled plasma mass spectrometry
IMAC	Immobilized metal affinity chromatography
IM-MS	Ion mobility mass spectrometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB medium	Lysogeny broth
M. xanthus	*Myxococcus xanthus*
MALDI MS	Matrix-assisted laser desorption/ionization mass spectrometry
MM	M9 minimal medium
MS	Mass spectrometry
MW	Molecular weight
NaF	Sodium fluoride
nESI-MS	Nano-electrospray ionisation mass spectrometry
N. europaea	*Nitrosomonas europaea*
Non-EncFtn	Non-encapsulin associated ferritin-like protein
OPD	*ortho*-phenylenediamine
P. furiosus	*Pyrococcus furiosus*
PDB	Protein Data Bank
Pfc-EncFtnS	$PFC\_05175_{1\text{-}99}$
pI	Isoelectric point
PMF	Peptide finger printing

R. rubrum	*Rhodospirillum rubrum*
RBS	Ribosome binding site
$RMSD_{C\alpha}$	Root-mean-square deviation of Cα positions
Rru-Enc	*R. rubrum* encapsulin or Rru_A0974
Rru-EncFtn-Enc	coexpressed Rru_A0973 and Rru_A0974
Rru-EncFtnSH	$Rru\_A0973_{1\text{-}96}$ with C-terminal hexahistidine tag
Rru-Enc-his	Rru_A0974 with C-terminal hexahistidine tag
Rru-Enc-SeMet	Rru_A0974 with SeMet labelled
SCU	Sucrose cushion ultracentrifugation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SeMet	Selenomethionine
SGU	Sucrose gradient ultracentrifugation
T. maritima	*Thermotoga maritima*

List of figures

List of tables

Chapter 1 Introduction

1.1 Why is iron important to cells?

Iron is the most abundant transition metal on earth and ubiquitous in almost all life forms (Theil & Raymond 1994). Iron is mainly involved in oxidoreductase to transfer electrons (Waldron et al. 2009). Iron can be bound by haem prosthetic groups of proteins, such as cytochrome oxidases (Svensson-Ek et al. 2002), catalases (Carpena et al. 2005) and peroxidases (Zámocký et al. 2015) that participate in the activation of molecular oxygen; haemoglobins and myoglobins that function as oxygen carriers (Ponka 1999); cytochromes involved in electron transport (Gursahani et al. 2008). Iron can also exist in proteins in a form of iron-sulphur clusters, such as ferredoxins, which function as electron carriers (Kümmerle et al. 2000). Furthermore, iron can be directly bound by protein side chains, including mononuclear non-haem iron enzymes; dinuclear non-haem iron enzymes such as ribonucleotide reductase; which participates in the early stages of DNA synthesis; as well as proteins engaged in iron storage (ferritins) and transport (transferrin) (Lill 2009; Crichton 2009). These iron-bound proteins are widely distributed across all Kingdoms except *Lactobacillus plantarum*, which has no iron containing enzymes (Archibald 1983). Iron is also pivotal in the host and pathogen interaction (Ratledge & Dover 2000; Johnson 2008).

Iron exists in cells in more than one oxidation state, with the two most common valences—Fe(II) and Fe(III). The iron valence is largely dependent on the redox potential and pH of the environment. The oxidation state of iron is very important for cells due to the drastically different solubility of two iron ions. Fe(II) is very soluble; for instance, the solubility of ferrous ammonium sulphate is 0.68 M in water at 20°C. However, Fe(II) is very easily to oxidise into Fe(III) in aerobic environment and thus generates oxyhydroxide minerals in the presence of water. These minerals have a very poor solubility; for example, $Fe(OH_2)^+$ has a solubility of 1.4×10^{-9} M at pH = 7 (Chipperfield & Ratledge 2000). Since the cellular iron requirement is around 10^{-4} M (Theil 2003), the poor solubility of Fe(III) in physiological conditions limits the bioavailability of iron for cells. It also means that aerobic organisms have more difficulty in acquiring iron than anaerobes, as the soluble Fe(II) is favoured in the

anaerobic environment. While iron is essential for many biological processes the presence of iron ions within a cell poses a number of problems for biomolecules due to the Fe(II)-induced Fenton reaction (Wardman & Candeias 1996).

1.2 Fenton chemistry

The evolution of organisms from anaerobes to aerobes has been partially influenced by the accumulation of di-oxygen in atmosphere. This accumulation was a direct result of cyanobacterial photosynthesis, which uses water as a hydrogen donor to produce sugars and di-oxygen. Fe(II) is stable under anaerobic and acidic conditions; however, it is readily oxidised by hydrogen peroxide, which is a by-product of aerobic respiration at physiological pH. This iron-catalysed hydrogen peroxide conversion to hydroxyl radicals was termed the Fenton reaction (Wardman & Candeias 1996; Fenton 1876) and proceeds as follows:

$$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + {}^{\bullet}OH + OH^{-}$$

The hydroxyl radicals ($^{\bullet}OH$) produced attack neighbouring molecules, causing endogenous DNA oxidation (Cadet et al. 1999), producing covalently bound protein aggregates (Davies 1987) and leading to lipid peroxidation (Ayala et al. 2014). This damage is mitigated by three major mechanisms in cells: compartmentalisation; antioxidant systems; and enzymes that directly degrade reactive oxygen species, such as iron containing-superoxide dismutase, superoxide reductase, and catalase (Riley 1994; Sheng et al. 2014). The limited Fe(III) solubility and Fenton chemistry-related toxicity requires strict iron homeostasis in almost all organisms. Perturbation of intracellular iron concentration can lead to cellular damage and severe health problems, such as the iron-overloading in hemochromatosis and multiple metabolic disorders, or iron deficiency in anaemia (MacKenzie et al. 2008).

1.3 How do organisms regulate iron metabolism?

1.3.1 Iron uptake

Cellular iron metabolism includes iron uptake, storage, mobilisation, and regulation processes (Aisen et al. 2001). The iron availability differs for marine, terrestrial and pathogenic microbes growing in different environment, which requires organisms to

take advantage of multiple iron uptake pathways (Sandy & Butler 2009). Iron is absorbed from soil or ocean through Fe(III) chelation by siderophores extracellularly-secreted by microbes or organic acids, phenols and phytosiderophores in plants (Colombo et al. 2014). In human 60% of the iron resides in haemoglobin, making it a main source for pathogens to capture iron from their hosts (Theil & Raymond 1994). Therefore, haem-mediated iron acquisition has also been utilised for pathogens in addition to siderophores (Braun 2001). For vertebrates, iron can be assimilated from dietary components by the mucosal cells of jejunum and transported to the blood stream, followed by chelation and transport by transferrins (Ratledge & Dover 2000).

1.3.2 Cellular iron storage

Although the haem-iron in haemoglobin accounts for the majority of iron present in the human body, haem itself does not act as an iron sink or storage system. Instead, ferritins are the primary iron storage proteins in most living cells, maintaining the intracellular iron availability in addition to extracellular iron sources (Theil 1987). Ferritins are a family of self-assembling protein cages with diameters of around 7–12 nm, enclosing a central cavity where iron can be reversibly stored as ferrihydrite (Fe_2O_3•H_2O). Iron-free ferritin is referred to as apoferritin; the maximum iron sequestered within the ferritin cage is around 4500 iron ions (Ford et al. 1984). This iron sequestration ability grants ferritin dual functions in providing a cellular iron reserve and protecting cells from Fe(II) toxicity (Zhang & Orner 2011). Most organisms possess multiple ferritin genes, highlighting their key roles in iron homeostasis (Bai et al. 2015). As ferritins play such an essential role in iron regulation, their structure and function will be discussed in more details below.

1.3.3 Iron release

When cytosolic iron is depleted iron is released by endogenous ferritin iron storage proteins. The iron release from ferritin mineral core *in vitro* involves the addition of electrons to Fe(III) as well as water/protons to the oxo/hydroxo ferritin bridges (Theil et al. 2000). How iron is mobilised and released from ferritins *in vivo* is still poorly understood, four potential mechanisms regarding iron release are suggested as follows: (1) the iron stored in ferritin cage is in an equilibrium with cytosolic iron (Laulhère et al. 1995); and transferrin facilitates the Fe(III) release as an iron scavenger (Ebrahimi

et al. 2012); (2) Fe(III) in the mineral core is reduced by biologically relevant reductants, such as NADH/FMN (Mertz & Theil 1983), before Fe(II) is released by chelators at the physiologically required rate (Watt & Jacobst 1988; Jones et al. 1978); (3) Fe(III) is released in eukaryotes through ferritin degradation within lysozymes (Kidane et al. 2006); (4) chaperones are involved in docking ferritin and directly removing Fe(III) (Carmona et al. 2013). Additionally, the haem groups in bacterioferritins was suggested to facilitate the reduction of Fe(III) mineral core and Fe(II) release via an electron transfer mechanism (Yasmin et al. 2011) involving ferredoxins which associate with bacterioferritin and possesses a [2Fe-2S] cluster. Ferredoxins are suggested to reduce the Fe(III) in bacterioferritin through electron transfer from cellular reducing agents such as NAD(P)H via the haem groups coordinated by the bacterioferritin (Weeratunga et al. 2009). It is also worth pointing out that iron scavenging by transferrin under physiological conditions is slower than the iron deposition in ferritin cages, making transferrin less possible as an iron release route. Lysozyme-dependent ferritin degradation may release iron in an uncontrollable way, which is not ideal in terms of isolating cells from Fe(II) toxicity. Overall, the iron release *in vivo* still needs more experimental evidence.

1.4 Ferritin family proteins

1.4.1 Ferritin classification

Ferritin was initially isolated from horse spleen in 1937 (Laufberger 1937); and the first crystal structure was determined in 1991 as a 24-meric protein cage with 432 octahedral symmetry (Lawson et al. 1991). The subsequent discoveries of ferritins in other organisms, such as mouse, human, horse (Rucker et al. 1996; Hempstead et al. 1997), soybean (Proudhon et al. 1996; Masuda et al. 2010a) and *Escherichia coli* (Hudson et al. 1993) revealed a similar architecture as the horse spleen ferritin, suggesting the importance of the conserved ferritin structure to its function. In 1998 the first structure of a mini-ferritin was determined from protein isolated from *E. coli* cells, known for its role in DNA protection during oxidative stress (DPS), as a 12-meric protein cage which suggested a distinct ferritin function (Grant et al. 1998).

Mammalian ferritins are heteropolymers consisting of two subunits: H-chain and L-chain. They share 56% amino acid sequence identity and are structurally homologous,

with a root-mean-square deviation of Cα positions ($RMSD_{Cα}$) of 0.486 Å for 153 atoms. H-chain ferritin is predominant in heart and has a molecular weight of around 21 kDa, with 182 amino acids. L-chain ferritin is about 19 kDa, with 174 amino acids and predominates in liver (Theil 1987). Mammalian ferritins isolated from different tissues vary in the ratios of H-chain/L-chain from 2:22 to 20:4, suggesting tissue-specific iron storing ability. (Levi et al. 1994; Wang et al. 2006). Apart from H-/L-chain ferritins, amphibians have an additional type M ferritin, which has been characterised in bullfrog (FgMF, PDB ID: 1MFR) (Ha et al. 1999). H-chain (Levi et al. 1988) or M-chain (Pereira et al. 1998) ferritin is responsible for the fast iron oxidation conferred by the ferroxidase centre (FOC) while L-chain ferritin does not possess ferroxidase centre (FOC) and is only involved in iron mineralisation (Wang et al. 2006; Levi et al. 1994).

There are three major classes of ferritins in microbes: bacterial ferritins (FTN), bacterioferritins (BFR), as well as DNA-binding protein from starved cells (DPS). They are able to catalyse the oxidisation of Fe(II) to Fe(III), and store iron in a mineral form in ferritin cages. But the chief role of DPS is not storing iron; instead, the ferrous-loaded *E. coli* DPS was found to protect DNA from oxidative damage in a long distance through two steps: ferrous was oxidised in DPS with DNA charge transport-dependent electron transfer, which was able to fill the guanine radical hole caused by the DNA oxidation and thus restore the DNA integrity (Arnold & Barton 2013). DPS, on top of ferroxidase activity, has also been found to possess peroxidase activity in *Microbacterium arborescens* (Pesek et al. 2011). Phosphate is suggested to be involved in the mammalian and bacterial ferritin cores and affect both iron deposition and release (de Silva et al. 1993; Watt et al. 1992; Watt et al. 2010).

1.4.2 Ferritin architecture

The quaternary structure of ferritin proteins is entirely critical to their function as iron stores (Crichton & Declercq 2010). The classical 24-meric ferritin fold consists of a four-helix bundle arranged in an up-down-down-up topology, followed by a short α5 helix that lies roughly perpendicular to the central axis of four-helix bundle, a long loop links α2 and α3 (**Figure 1.4.1 A–C**).

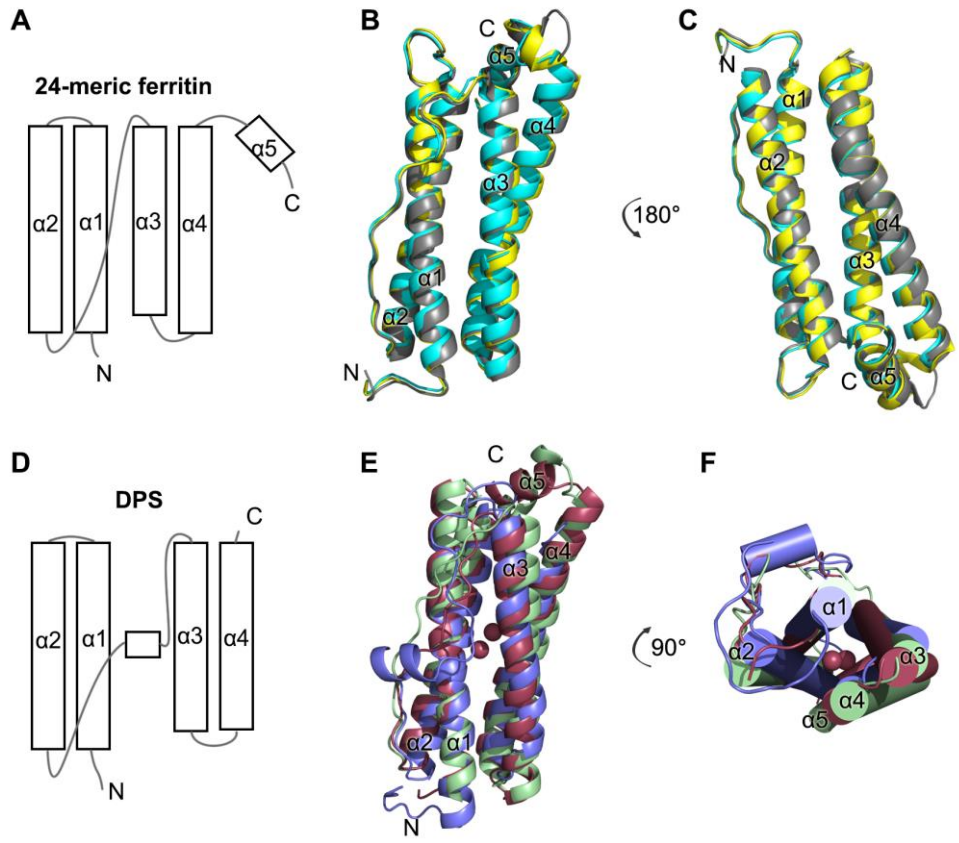


Figure 1.4.1 The four-alpha helical ferritin fold.

The ferritin fold is characterised by a four-alpha-helix bundle coordinating di-nuclear metal ions, followed by a short α5 in the C-terminus. (A) The arrangement of 24-meric ferritin fold is shown. (B) The front view of superimposed human H-chain ferritin (yellow cartoons, PDB ID: 2FHA) (Hempstead et al. 1997), M-type ferritin (cyan cartoons, PDB ID: 1R03 (Langlois d'Estaintot et al. 2004) and L-chain ferritin (grey cartoons, PDB ID: 2FG4) (Wang et al. 2006). (C) Back view of (B). (D) The arrangement of 12-meric DPS fold is shown. (E) The front view of superimposed E. coli bacterial ferritin (green cartoons, PDB ID: 1EUM) (Bou-Abdallah et al. 2014), Azotobacter vinelandii bacterioferritin shown in raspberry cartoons with di-iron bound in spheres (PDB ID: 2FL0) (Swartz et al. 2006), and E. coli DPS (slate cartoons , PDB ID: 1L8H). (F) The top view of (E) with alpha helices shown in cylindrical mode. Notably, DPS does not contain α5; instead, it possesses a short alpha-helical region in the loop connecting α2 and α3.

The four-helix bundle coordinates iron ions in a di-nuclear ferroxidase centre, which is present in all classical ferritins except the L-chain ferritins (Levi et al. 1994). Twenty-four monomers of classical ferritin assemble into a 12-faced polyhedron with around 480 kDa molecular weight; this assembly has a 432 octahedral symmetry. This typical 24-meric spherical structure, which has an outer diameter of roughly 12 nm and an inner diameter of 8 nm, is common to eukaryotic ferritins, FTNs and BFRs (**Figure 1.4.2 A/B**).

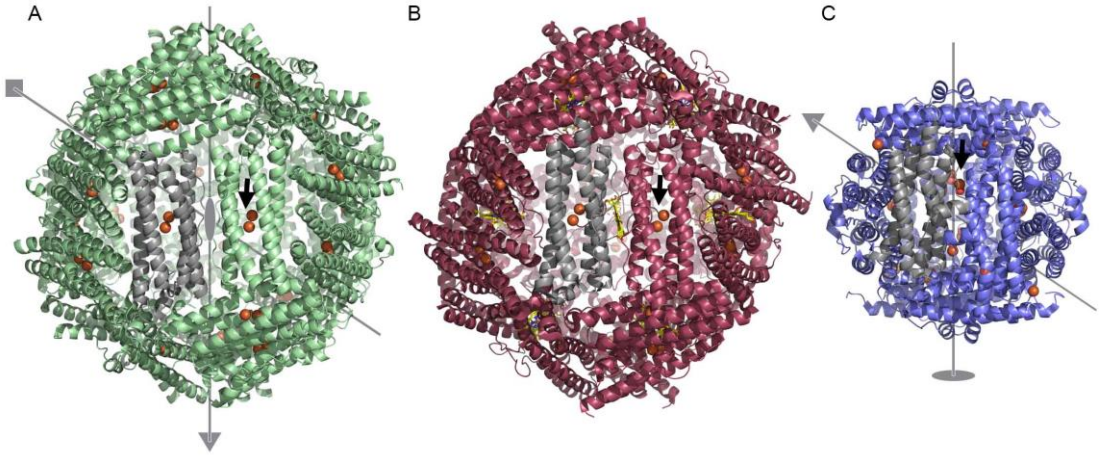


Figure 1.4.2 Structure of microbial ferritins.

The quaternary structures of FTN, BFR and DPS are shown in cartoon mode; a single monomer is coloured grey and bound iron ions in ferroxidase active sites are shown as orange spheres. (A) Structure of FTN from *Pseudonitzchia multiseries* (PDB ID: 4IWJ) (Pfaffen et al. 2013). Grey lines and polygons depict the relationship between the symmetry axes of the ferritin cage: 2-fold, ellipse; 3-fold, triangle; 4-fold, square. (B) Structure of BFR from *Desulfovibrio desulfuricans* (PDB ID: 1NF4) (Macedo et al. 2003). The bound *b*-type haem groups are shown as yellow sticks between the protein monomers; due to the co-ordination of this flexible group the 2-fold axes between monomers are not true symmetry axes. BFR is shown in the same orientation as FTN. (C) Structure of DPS from *Microbacterium arborescens* (PDB ID: 2YJK) (Pesek et al. 2011). DPS is a dodecamer and has 2- and 3-fold symmetry axes; the relationship between these is illustrated with grey lines and a triangle and ellipse. This figure is adapted from (He & Marles-Wright 2015).

Classical ferritins from various organisms and tissues differ in iron storage ability, suggesting that different types of ferritins have evolved specialised functions in their hosts and host tissues in multicellular organisms. Mammalian ferritin is able to store up to 4,500 iron ions in its cage (Ford et al. 1984). DPS, also called mini-ferritin, consists of 12 subunits arranged in a cage with 23 point group symmetry, with 9.5 nm outer diameter and 5 nm inner diameter; the central cavity can store up to 500 iron atoms (Zeth 2012). DPS differs from the classical 24-meric ferritin fold in four aspects: (1) α5 is missing in DPS; (2) DPS contains a short alpha helical region in the α2-α3 loop; (3) the FOC of DPS lies at the two-fold symmetry interface instead of the centre of a monomer; (4) the α3 and α4 in DPS are shorter than the 24-meric ferritin counterparts (**Figure 1.4.1 D–F and Figure 1.4.2 C**). Several structures of L-chain ferritin have been determined including horse (Gallois et al. 1997), mouse (Granier et al. 2003) and human (Wang et al. 2006) variants. These show the FOC is missing; thus, rather than functioning in iron oxidation, the L-chain ferritin is suggested to facilitate iron mineralisation (Levi et al. 1994). Both FTN and BFR resemble H-chain ferritin

structures, but BFR coordinates up to 12 *b*-type haem groups at the 2-fold symmetry axes between the subunits (**Figure 1.4.2 B**). The iron in these haem groups is coordinated by conserved methionine residues and it is suggested that the presence of haem helps release iron from iron minerals inside the BFR cage by facilitating electron transfer across the ferritin cage (Yasmin et al. 2011).

1.4.3 Ferritin assembly/disassembly

Some ferritin cages are very stable and are able to withstand temperatures up to 85°C, and the presence of chaotropic agents, such as 6 M urea, or guanidine, when the pH is above 3.5 (Listowsky et al. 1972). The *in vitro* assembly and disassembly of ferritins have been shown to be pH-dependent. The horse spleen apoferritin is found out to disassemble when pH is lower than 0.8 while the ferrihydrite core starts to aggregate under pH 2.1. However, the horse spleen apoferritin can only be reconstructed to 20-mer when pH changes from 3.24–1.96 to neutral conditions. It is suggested to be related to the specific intersubunit interaction and requires other molecules to assist in the reassembly *in vivo* (Kim et al. 2011). The self-assembly of ferritin relies on the intersubunit electrostatic interactions between two loops which connect α2 and α3 perpendicular to the 2-fold symmetry axis, as shown in **Figure 1.4.2 A** (Bernacchioni et al. 2014). In *E. coli* BFR ferritin appears to assemble from a stable intermediate—dimer species, as shown by alanine scanning mutagenesis of 2-fold, 3-fold or 4-fold symmetry interface residues, of which the formation of 24-meric BFR cage was abolished, but dimers remained (Zhang et al. 2010). The haem groups are suggested to be not essential for the assembly of bacterioferritin (Andrews et al. 1995). It was discovered in a bacterial ferritin from the archaeon *Pyrococcus furiosus* that the formation of the 24-meric ferritin cage is a prerequisite for fulfilling Fe(II) oxidation (Ebrahimi et al. 2015a). The assembly of ferritins has been engineered to be controllable by modifying the C_2 interface residues to be specifically bound by copper ion at neutral pH, followed by disrupting the other C_2 interfacial interaction, so that the engineered ferritin will not assemble only if Cu(II) is present (Huard et al. 2013).

1.4.4 Iron binding sites in classical ferritins

It is not easy to determine the iron binding in ferritins due to its relatively low-affinity binding at sites (K_d around 10^{-5} M) other than the ferroxidase centre (Bertini et al. 2012). The transient iron binding can be captured by slowing down the iron movement under anaerobic conditions or using competitors such as Mg(II) or Zn(II) (Ebrahimi et al. 2010). Four iron binding sites have been identified in HuHF based on time-dependent iron soaking in anaerobic crystals (**Figure 1.4.3**), including di-iron (Fe_A and Fe_B) bound in the FOC, another two iron ions (Fe_C and Fe_D) bound in site C next to the FOC, two iron ions coordinated in the three-fold symmetry channel, as well as an iron ion found in the four-fold symmetry pore (Pozzi et al. 2015). The ferritin chemistry includes iron entry, oxidation, translocation, mineralisation and exit, which are separated in space and time (Theil et al. 2000).

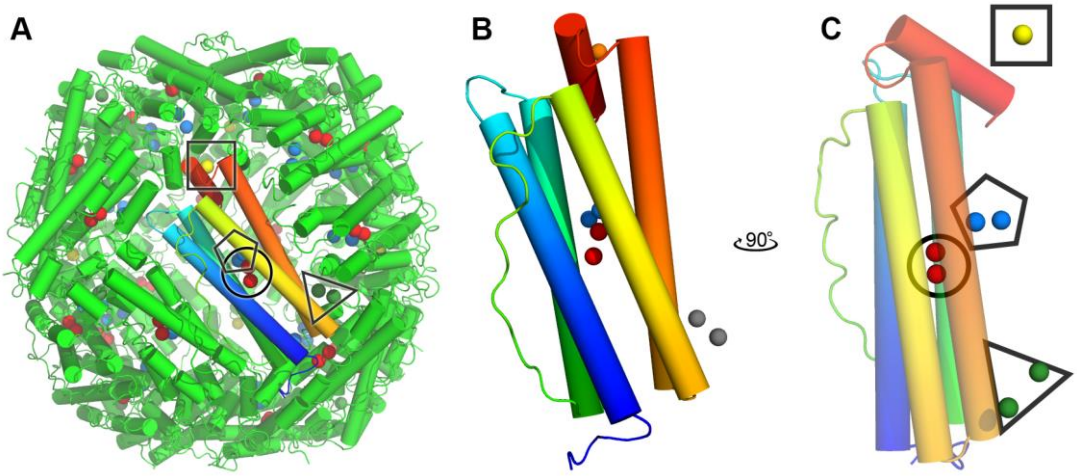


Figure 1.4.3 Iron binding sites in human H-chain ferritin.

The structure of human H-chain ferritin (HuHF, PDB ID: 4YKH) is shown in cartoon mode with iron shown in spheres (Pozzi et al. 2015). The iron binding in different sites is highlighted in different colours. (A) 24-meric HuHF is shown in green cartoon with a monomer highlighted in rainbow colour. (B) The monomeric HuHF. (C) Rotated view of (B) by 90° displays the four iron binding positions. Di-iron (Fe_A and Fe_B) bound in the FOC is shown in red spheres and highlighted by a circle. A pentagon underscores two iron ions (blue spheres) coordinated in the Fe_C site close to the FOC. A triangle highlights two iron ions (green spheres) bound in the three-fold symmetry channel. A square highlights an iron ion (yellow sphere) bound in the four-fold symmetry channel of HuHF.

1.4.4.1 Ferroxidase catalytic site

All classical ferritins except the L-chain ferritin possess a highly conserved FOC where two iron ions (Fe_A and Fe_B) are coordinated by at least three glutamic acid residues and one histidine residue, with metal coordination distances of 2.4 ± 0.3 Å (Ebrahimi et al. 2015b). The distance between the two iron ions is 3.5 ± 0.2 Å (**Figure 1.4.4 A/B**). In the FOC of HuHF, Fe_A in site A is coordinated by bidentate E27, monodentate E62 and H65; Fe_B in site B is bound by bidentate E107, monodentate E62, and Y34, which is found 4.6 Å away from Fe_B (**Figure 1.4.4 A/B**). The tyrosine residue is highly conserved in classical ferritins and is suggested to facilitate Fe(II) oxidation as a single-electron molecular capacitor (Ebrahimi, Hagedoorn & Wilfred R Hagen 2013). Site A is more strongly conserved than site B in terms of the residues involved in iron coordination (**Figure 1.4.4 A/B**), which suggests differential binding affinities of the iron ions in the FOC (Ebrahimi et al. 2015b).

The FOC functions as a catalytic site where Fe(II) can be oxidised by di-oxygen. This is supported by the abolished Fe(II) oxidation in the E62A or H65A mutants of recombinant HuHF (Sun et al. 1993). In the vicinity of the FOC there is a third iron (Fe_C) binding site (site C), which is located at a distance of 6.2-10 Å to Fe_A/Fe_B and towards the inner face of the ferritin cage (**Figure 1.4.3 C** and **Figure 1.4.4 C/D**). Among four aligned ferritins [HuHF, frog M-chain ferritin (FgMF), PfFtn and EcBfr], site C displays higher variability with different residues, conformations and numbers of iron ion than the FOC (**Figure 1.4.4 C/D**). Notably, the FOC is not always fully occupied with iron, which largely depends on crystallisation conditions such as pH; however, site C is either partially occupied or vacant (Ebrahimi et al. 2015b). This site C is highly conserved in bacterioferritins (Ebrahimi et al. 2015b); the mutation of conserved residues in site C of HuHF (Masuda et al. 2010b), bullfrog M-type ferritin (Behera & Theil 2014) or PfFtn (Ebrahimi et al. 2012) all lead to an abrogated iron oxidation, suggesting that site C is involved in the translocation of Fe(II)/Fe(III) into/out of the FOC.

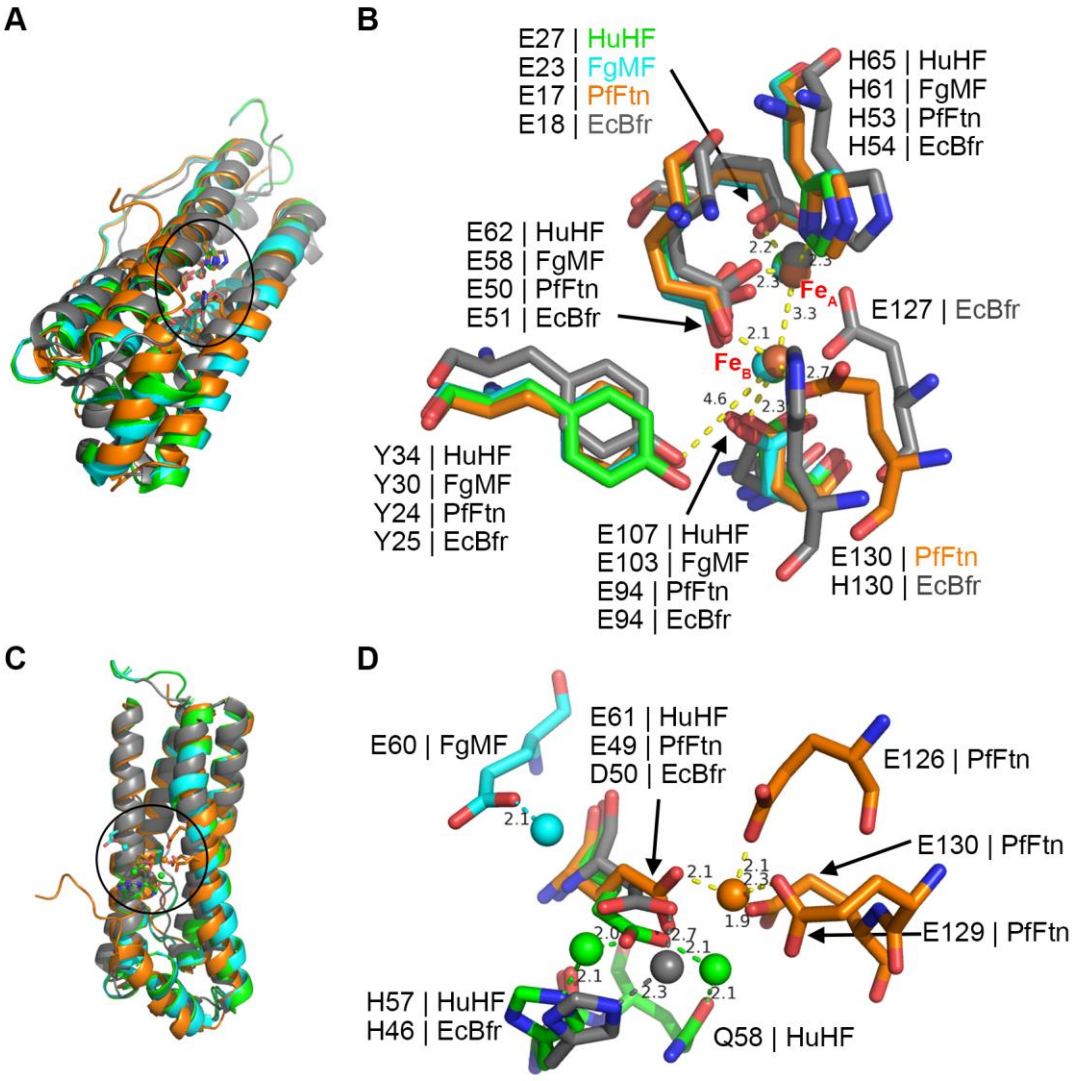


Figure 1.4.4 Ferroxidase centre and gateway site of classical ferritins.

Monomers of HuHF (PDB ID: 4YKH), FgMF (PDB ID: 4LQN), PfFtn (PDB ID: 2JD7) and EcBfr (PDB ID: 3E1M) were aligned in PyMOL with an overall $RMSD_{C\alpha}$ < 0.7 Å. These are shown in green, cyan, orange and grey cartoons, respectively. (A) The di-iron centre is shown highlighted in an ellipse. (B) Enlargement of the region highlighted in (A) showing the highly-conserved FOC (Fe_A-Fe_B), with labelled distances, in Ångströms, between di-iron and their coordinating residues of PfFtn. Iron binding residues are shown as sticks; Fe_A and Fe_B are shown as spheres and coloured according to their coordinating residues. (C) View of the aligned monomeric ferritins showing the Fe_C site highlighted with an ellipse. (D) Enlargement of the region highlighted in (C) showing the highly-variable Fe_C site (gateway site) of aligned ferritins. Residues and coordinated metal ions are shown as in (B).

1.4.4.2 Putative iron entry and exit sites in classical ferritins

When cells have an abundance of Fe(II) it must traverse the ferritin shell from the cellular environment prior to Fe(II) oxidation in the FOC and deposition in ferritin cage; conversely, while cells are iron-depleted, ferritin has to release iron back to the cytoplasm. Understanding how ions/electrons transport into and out of the ferritin cage will provide insights into how cells regulate their iron concentrations. The iron entry and exit from ferritin has been studied using computational calculations of electrostatic gradients (Douglas & Ripoll 1998) and MD simulations (Laghaei et al. 2014), combined with biochemical assays (Theil et al. 2008), biophysical methods including isothermal titration calorimetry, and X-ray crystallography of relevant mutants of ferritin channels (Bou-Abdallah et al. 2003). In spite of the wealth of data and literature in the field, iron entry and release from ferritins is still not fully understood.

Both the phase and oxidation state of iron change within the ferritin nanocage during its catalytic cycle. It is generally agreed that the pores at the three-fold symmetry axes of the classical ferritins are the major entry site for Fe(II) while the four-fold symmetry channels are the primary Fe(II) exit route out of the non-haem ferritin cage (X Yang et al. 2000). The 3-fold or 4-fold channel residues appear to be not conserved among ferritin proteins (Ebrahimi et al. 2015b). An example of iron translocation into and out of HuHF is demonstrated in **Figure 1.4.5** (Pozzi et al. 2015). The pore size of both 3-fold and 4-fold symmetry channel measured in the crystal structure of HuHF is about 4.4–4.6 Å. The 3-fold symmetry pore is lined with negatively-charged residues (E134 and D131) and is shaped as a hydrophilic funnel, which guides Fe(II) or other divalent cations to traverse the ferritin shell. The hexahydrated iron ions binding in the HuHF 3-fold symmetry pore suggest that water may diffuse into ferritin from this 3-fold channel when hydrating Fe(II) (**Figure 1.4.5 C**). It has also been observed in horse L-chain ferritin that the electrostatic gradient in 3-fold channel impedes the entry of anions but does not strictly limit the entry of monovalent cations via diffusion unless divalent cations are present (Takahashi & Kuyucak 2003). In another experiment, cations such as sodium or potassium are demonstrated to compete with Fe(II) for the negatively-charged 3-fold symmetry channel. In anions only phosphate and some tetrahedral oxo-anion analogues are able to stimulate the iron loading into ferritin,

while sulphate decreases the iron loading to a higher extent than cations and halides (Cutler et al. 2005). The specificity of anion transport into ferritin is suggested to be driven by the solubility of ferritin and anion complex. That is to say, if the anion prefers to form an insoluble substance with iron, this complex settles inside ferritin; if the anion tends to form soluble complex with iron, then iron hydroxide forms (Cutler et al. 2005). However, anions are repelled by the negatively-charged residues lining the 3-fold pores, as supported by electrostatic calculations (Douglas & Ripoll 1998), and diffusion studies using nitroxide radical spin probes in conjunction with electron magnetic resonance spectroscopy (EPR) (Yang & Chasteen 1996; X Yang et al. 2000). This suggests that the entry of anions such as phosphate ions may rely on other process. The selectivity of divalent cations through 3-fold channel is also not clear. On top of charge selectivity, the pore size of the 3-fold symmetry channel also controls the substance entry. The EPR study reveals the entry of positively-charged spin probe with around 7–9 Å cross section through 3-fold symmetry channel, in which the spin probe serves as a model for similar sized reductants and chelators. The results suggest that the ferritin surface pore with around 4 Å size in the crystal structure may be somewhat flexible and adopt a breathing mode to accommodate the transport of molecules across the ferritin cage (X Yang et al. 2000; Yang & Chasteen 1996).

The hydrophobic four-fold channel in human H-chain ferritin coordinates iron via H173 of four chains, which are located at inner side of ferritin shell (**Figure 1.4.5 C/D**). The large electrostatic potential gradient at the 4-fold symmetry channel presents an electric field from the inner cavity to the outer environment, suggesting that this channel serves to expel cations from ferritin cage (Douglas & Ripoll 1998). In *Helicobacter pylori* FTN structure the 4-fold channel can be the entry route of Fe(II) when more than 500 Fe(II) per 24-mer is loaded aerobically (Cho et al. 2009). The second possible role of 4-fold channels is to provide an exit for excess protons which are generated during the iron mineralisation process (Takahashi & Kuyucak 2003).

It is not clear if the *in vivo* Fe(II) entry route remains the same as is shown *in vitro* experiments and if the iron entry routes are conserved among different members of the ferritin family from different organisms. An alignment of classical ferritin sequences (**Figure 1.4.6**) shows that the E134 and D131 of HuHF in the 3-fold channel and H173

in the 4-fold pore are only conserved in FgMF. Thus, it remains uncertain if the iron translocation model of HuHF holds for FTN and BFR. It has been proposed that there are another three putative Fe(II) entry routes in FTN and BFR apart from the 3-fold channel: (1) FOC channel, where Fe(II) directly moves to the FOC from the outer face of the nanocage and then to site C; (2) B-pore (intersubunit gap) where Fe(II) moves through and then translocates to the FOC, or site C; this has been seen as the principal iron entry route for *E. coli* bacterioferritin (Wong et al. 2014); (3) 4-fold channel where Fe(II) traverses the protein shell and moves to FOC or site C (Ebrahimi et al. 2015b).

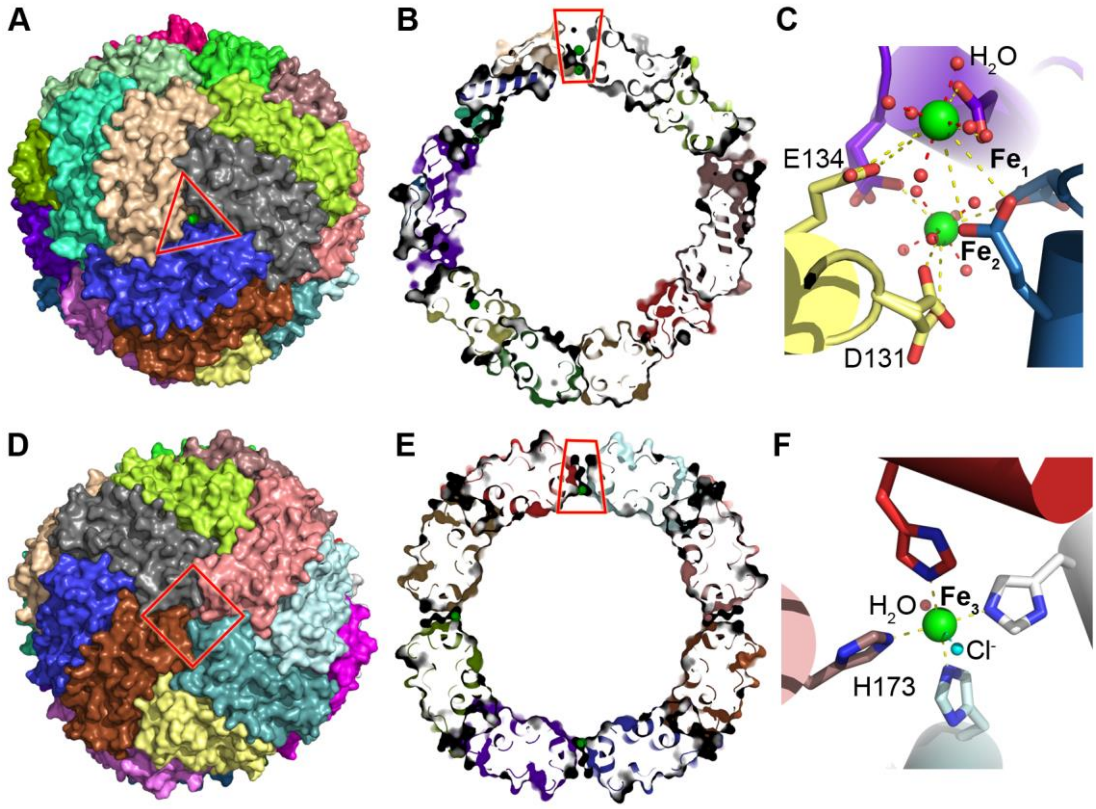


Figure 1.4.5 Iron entry/exit site of human H-chain ferritin.

HuHF (PDB ID: 4YKH) is shown in surface mode in (A, B, D, E). (A) A red triangle highlights the 3-fold symmetry pore, which is a putative Fe(II) entry site. (B) A 90° rotation and cross-section view of (A) is shown, in which the two iron ions (Fe_1 and Fe_2 in green spheres) bound in 3-fold channel are highlighted. (C) The coordination environment of the two iron ions in the entry site is shown, with amino acids show as sticks, iron shown as green spheres, and water as red spheres. Hexahydrated Fe_1 is coordinated by bidentate E134 from three chains at a distance of 3.9–4.2 Å. Hexahydrated Fe_2 is bound by monodentate D134 from three chains in 2.6 or 3.6 Å as an alternate conformation. This differential iron binding leads to a funnel-like three-fold channel. (D) A square highlights the 4-fold symmetry pore, which is a putative iron exit site. (E) A 90° rotation and cross-section view of (D) is shown, in which an iron ion (Fe_3) in green sphere bound in 4-fold channel is highlighted. (F) Fe_3 in the 4-fold symmetry pore is bound by H173 from four chains as well as a water and chlorine ion in a distance of 2.2 Å.

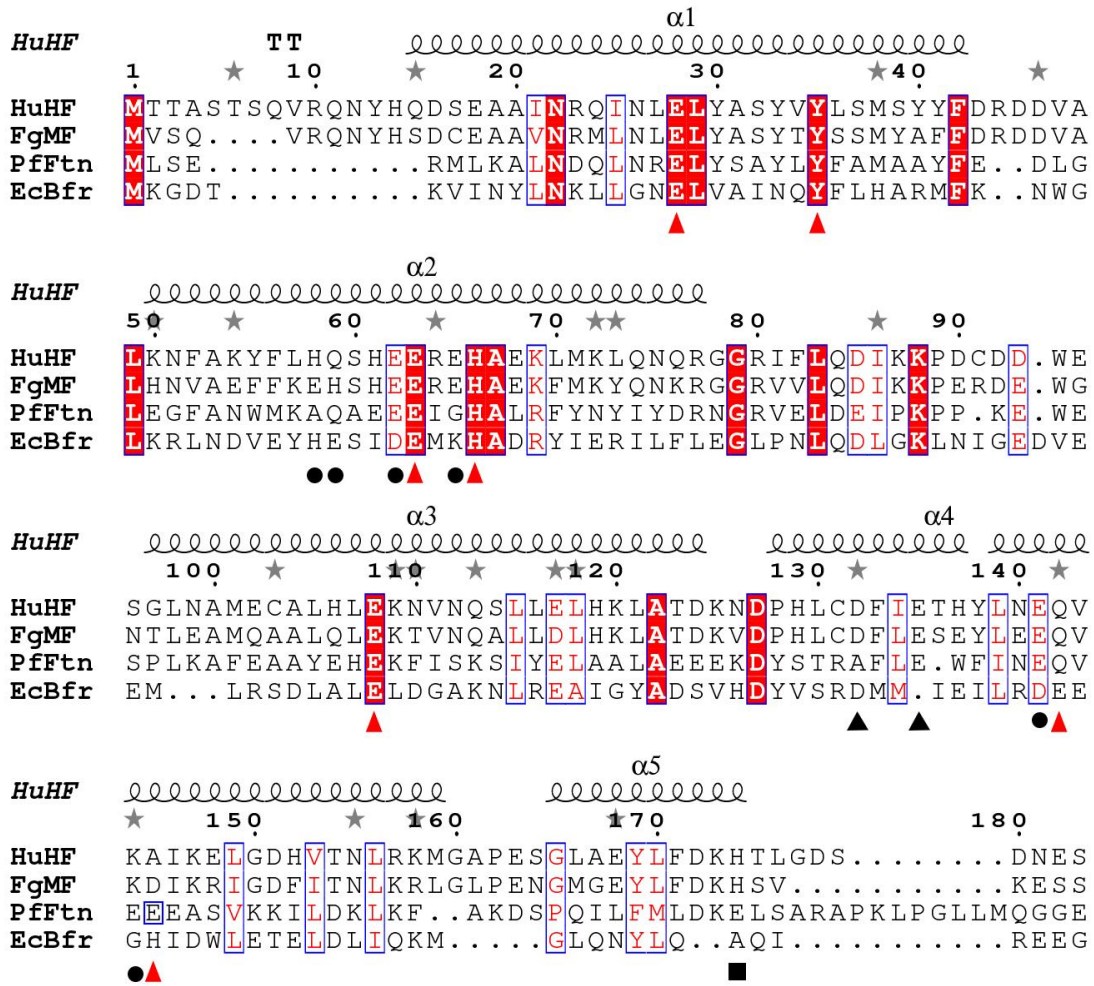


Figure 1.4.6 Alignment of iron binding sites of classical ferritins.

Human H-chain ferritin (HuHF, PDB ID: 4YKH), frog M-chain ferritin (FgMF, PDB ID: 4LQN), bacterial ferritin from *P. furiosus* (PfFtn, PDB ID: 2JD7), and bacterioferritin from *E. coli* (EcBfr, PDB ID: 3E1M) were aligned using Clustal W (Larkin et al. 2007; McWilliam et al. 2013). The sequence alignment was edited and displayed in Espript 3.0 (Robert & Gouet 2014) with red triangles underscoring FOC residues and black circles highlighting residues involved in binding Fe_C. Two black triangles highlight the residues binding two iron ions separately in the three-fold channel of HuHF. A black square highlights the residue binding iron in the four-fold symmetry pore of HuHF.

1.4.5 Ferritin chemistry

1.4.5.1 Ferroxidase reaction

The ferritin reaction stoichiometry has been studied in HuHF (Zhao et al. 2001), PfFtn (Ebrahimi et al. 2009), *E. coli* FTN (EcFtnA) (Bou-Abdallah et al. 2014) and *E. coli* BFR (EcBfr) (Xiaoke Yang et al. 2000) by measuring oxygen uptake and conventional and stopped-flow absorption spectrometry after aerobic ferritin titration with various amount of Fe(II). Sometimes catalase or horseradish peroxidase is added to the

ferroxidase assay to detect the formation of H_2O_2 (Zhao et al. 2001). In heterologously expressed eukaryotic ferritins the presence of L-chain does not affect the iron oxidation stoichiometry; instead, the L-chain facilitates the iron deposition in ferritin cage (Yang et al. 1998). Two possible equations for the iron oxidation catalysis were proposed as follows (Ebrahimi et al. 2015b):

$$2Fe^{2+} + O_2 + 2H^+ \rightarrow 2Fe^{3+} + H_2O_2 \quad (1)$$
$$3Fe^{2+} + O_2 + 4H^+ + e^- \rightarrow 3Fe^{3+} + 2H_2O \quad (2)$$

When two Fe(II) per FOC are loaded to non-haem classical ferritins, half Fe(II) is suggested to be oxidised according to equation (1); another half is oxidised via reaction (2). The third Fe(II) in reaction (2) could be oxidised in site C by O_2 or hydrogen peroxide (H_2O_2) produced in (1). It has been suggested as a universal mechanism for ferroxidase reactions in non-haem classical ferritins of different organisms (Ebrahimi et al. 2015b). Therefore, when more than two Fe(II) per FOC are loaded to HuHF, PfFtn or EcFtnA, the stoichiometry of Fe(II) to O_2 increases from 2–3 to about 3.4–3.8. In HuHF and EcFtnA it has been proposed that the iron oxidation occurs in the FOC first; once the Fe(III) mineral core forms, the iron oxidation is catalysed on the mineral surface. However, in PfFtn the Fe(III) mineral core is not involved in Fe(II) oxidation. Instead, when the FOC and site C are both saturated by Fe(II), more Fe(II) oxidation occurs in the site C; together with some Fe(II) oxidation by hydrogen peroxide generated in reaction (1), the total stoichiometry of Fe(II):O_2 will be roughly 4:1 (Ebrahimi et al. 2015b).

A blue/purple intermediate of the ferritin reaction forms within a millisecond and decays in less than 1 second; this is always observed in eukaryotic, bacterial and archaeal ferritins via monitoring the absorbance at 500–800 nm (Pereira et al. 1998). The molecular structure of blue intermediates occurring in HuHF (Bou-abdallah et al. 2002) and FgMF (Pereira et al. 1998) have been identified as μ-peroxodiferric species (DFP) using Mössbauer spectroscopy and resonance Raman spectroscopy. The formation/decay of the blue intermediate (A_{650}) has been used to monitor how mutations in the ferritin iron entry route affect the ferroxidase activity (Behera & Theil 2014). Taken together, the ferroxidase reaction in ferritins involves Fe(II) binding to the catalytic site, followed by O_2 binding and production of a blue intermediate; this

μ-peroxodiferric complex decays transiently to diferric oxo mineral precursors in transit to the cavity (Tosha et al. 2010). The composition of the Fe(III) mineral core in horse spleen ferritin has been identified as a mixture of magnetite, hematite and ferrihydrite using techniques such as transmission electron microscopy, X-ray absorption near edge spectroscopy and electron energy-loss spectroscopy, small-angle X-ray scattering and SQUID magnetic measurements (Galvez et al. 2008). Phosphate has also been shown to facilitate the iron incorporation into ferritin (Polanams et al. 2005); in *P. furiosus* FTN, phosphate stimulates the displacement of Fe(III) by Fe(II) in the FOC; however, phosphate does not alter the iron oxidation mechanism as the formation of blue intermediates is not affected by the presence of phosphate (Ebrahimi, Hagedoorn & Wilfred R. Hagen 2013). The native iron mineral core presents Fe: P ≈ 1:1 in *A. vinelandii* ferritin, while Fe: P ≈ 8:1 in horse spleen ferritin (Rohrer et al. 1990). The presence of phosphate in a reconstituted ferritin alters the mineral core from crystallinity to an amorphous state, as the phosphate layer forms on the mineral core surface (Trefry et al. 1978).

1.4.5.2 Iron storage mechanism of classical ferritins

The asymmetric FOC of non-haem 24-meric ferritins appear to be related to a proposed iron storage mechanism, of which Fe(III) mineralisation/deposition in the cavity of a bacterial ferritin from *P. furiosus* (PfFtn), as well as human H-chain ferritin (HuHF), is driven by the sequential displacement of Fe(III) by Fe(II) in the FOC (Ebrahimi et al. 2012). In *E. coli* bacterioferritin (EcBfr, PDB ID: 3E1M), the FOC consists of two additional residues: a bidentate E127 coordinating Fe_A and Fe_B; and H130 which binds Fe_B, making the FOC of EcBfr more symmetrical. It suggests that bacterioferritin may have a different iron storage mechanism due to the different di-iron coordination environment (Jameson et al. 2002). In EcBfr the di-Fe(III) bound by the FOC is suggested to work as a catalytic cofactor site. Briefly, when the mineral core is absent in the BFR cage, Fe(II) ions are oxidised in the site C along with the reduction of oxygen to hydrogen peroxide in the di-Fe(III)-FOC through an electron-transfer mechanism. When there is mineral core formed in the cavity, the Fe(II) oxidation occurs on the surface of the mineral core and the electrons are transported to the di-Fe(III)-FOC via the Fe(II) bound in the site C, as site C locates between FOC and the

ferritin cavity (Crow et al. 2009). This complicated iron storage pathway is poised based on the slower iron oxidation observed when Fe(III) is present in the FOC. However, this observation has also been explained using the displacement model as Fe(II) substitutes Fe(III) in the FOC firstly before the Fe(II) oxidation (Ebrahimi et al. 2015b). In BFRs from *Pseudomonas aeruginosa* (Weeratunga et al. 2010), *Azotobacter vinelandii* (Swartz et al. 2006) and *Desulfovibrio desulfuricans* (Macedo et al. 2003), a common conformational change of the histidine residue in the FOC has been observed, suggesting a gating mechanism for iron translocation from the FOC to the cavity.

1.4.6 Phylogenetic tree of ferritin family proteins

The 24-meric ferritins including eukaryotic H-/M-/L-chain ferritins, bacterial ferritin and bacterioferritin, as well as 12-meric ferritins share the archetypal ferritin fold, in which the α1-α2 helical pair of the 24-meric ferritins is thought to be evolutionarily related to the α3-α4, as demonstrated by the alignment of two alpha-helical pairs in HuHF (**Figure 1.4.7 A**).

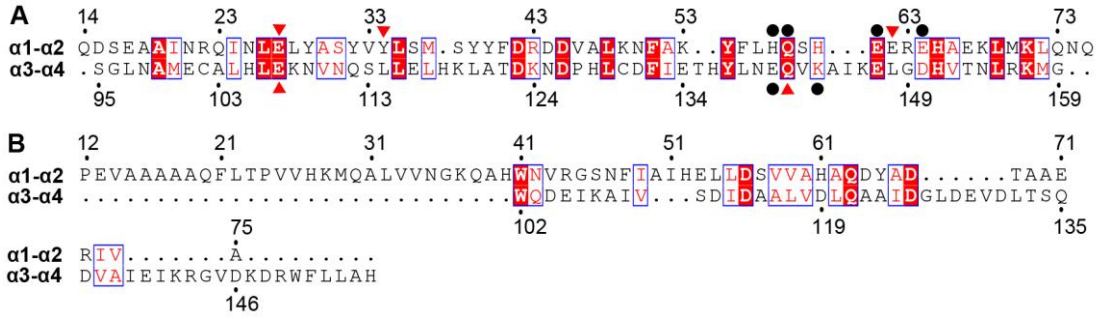


Figure 1.4.7 The alpha-helical pair alignments of human H-chain ferritin and DPS.

The α1-α2 helical pair of human H-chain ferritin (HuHF, PDB ID:4YKH) comprises residues 14−76 while the α3-α4 helical pair comprises residues 95−159 (Pozzi et al. 2015). The α1-α2 helical pair of *M. arborescens* DPS (PDB ID: 2YJK) contains residues 12−75, while the α3-α4 helical pair contains residues 102−155 (Pesek et al. 2011). The two helical pairs were separately aligned by Clustal Omega (McWilliam et al. 2013; Sievers et al. 2011), followed by editing with Esprit 3.0 (Robert & Gouet 2014). (A) The alignment of the two alpha-helical pairs of HuHF. The red triangles highlight the FOC residues (Fe_A and Fe_B site); and the black circles underscore the Fe_C site. (B) The alignment of the two alpha-helical pairs of DPS.

Although the protein sequence similarity of α1-α2 and α3-α4 is only 17%, the iron catalytic sites in the two helical pairs are well aligned. However, because the metal binding site of DPS is not in the centre of the ferritin fold, its α1-α2 and α3-α4 displays

relatively low sequence identity, suggesting that the two alpha-helical pairs in DPS may be evolutionarily distant from the classical ferritins (**Figure 1.4.7 B**).

The 24-meric ferritins were proposed to evolve from ancient rubrerythrins (Rr) or erythrins (Er) due to an internal duplication of the alpha-helical pair; and rubrerythrins were considered more homologous to the common ancestor than ferritins (FTN/BFR/DPS) (Andrews 2010). Rr protein is generally found in obligate anaerobes, and functions as a peroxidase in *Pyrococcus furiosus* (Weinberg & Jenney 2004) or a ferroxidase from *Desulfovibrio vulgaris* (Bonomi et al. 1996) that protects cells against redox stress. Rr contains two domains: an N-terminal Er domain with four-helix bundle coordinating a di-iron centre (~120 residues); and a rubredoxin (Rx) domain with a Fe-S_4 site at the C-terminus (~50 residues); both Er and Rx domains are required for the ferroxidase function (DeMaré et al. 1996). The Er domain structurally resembles ferritins regarding the ferritin fold as well as the conserved di-iron cluster. Therefore, S. Andrews suggested that entire ferritin families are likely to evolve from a simplistic two alpha-helices (i.e. α1-α2) without the loop linking α2 and α3; through gene duplication that a homo-dimeric four-helix bundle forms with anti-parallel arrangement (Andrews 2010).

1.4.6.1 Ferritin-like proteins with or without encapsulin association

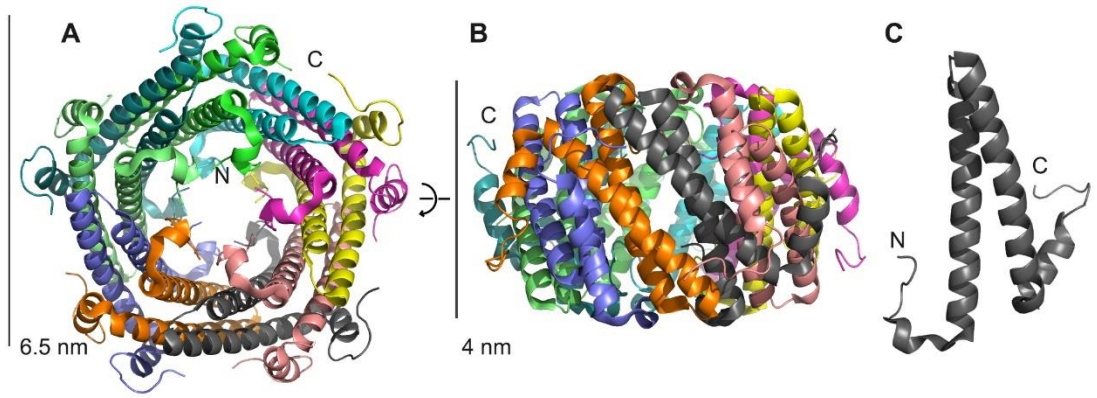


Figure 1.4.8 Structure of ferritin-like protein NE0167 from *Nitrosomonas europaea*.

The structure of NE0167 (PDB ID: 3K6C) is shown in cartoon mode and coloured by chain. The N- and C-termini have been highlighted, of which the N-termini of the structure cap the central channel while the C-termini point towards the outer surface. (A) The decamer ring-like structure has a diameter of 6.5 nm. (B) The side view of (A) possesses a height of 4 nm. (C) The monomer consists of two long alpha helices followed by a short alpha helix.

A ferritin-like protein with only two putative alpha helices is initially identified in *Nitrosomonas europaea* with a ring-like decamer structure deposited to the PDB (KEGG ID: NE0167, PDB ID: 3K6C) by the Midwest Centre for Structural Genomics (**Figure 1.4.8**). This decameric arrangement is radically different from all known ferritins (Chang et al. n.d.). NE0167 contains only 95 amino acids that constitute the two main alpha helices(α1-α2) followed by a short α3 helix. It is unclear what the physiological role of this ferritin-like protein is in its native host. A search of sequence databases for homologues of NE0167 identifies two classes of this family of proteins: the ferritin-like proteins and the ferritin-like protein with another protein associated in a conserved operon. The ferritin-like proteins are only found in microbes. In the latter case, the encoding gene is located upstream of a gene that encodes a protein structurally related to the gp5 capsid protein of the HK97 bacteriophage; this protein is called 'encapsulin' (Akita et al. 2007). This family of the ferritin-like proteins is termed the encapsulin associated ferritin-like proteins (**EncFtn**). EncFtn is structurally located inside the encapsulin shell through the interaction of a short C-terminal localisation sequence on the EncFtn with the interior surface of the encapsulin shell (Sutter et al. 2008). In some microbes, such as *N. europaea*, the ferritin-like protein does not have encapsulin associated in the genome and does not possess the C-terminal localisation sequence, these are thus termed the non-encapsulin associated ferritin-like protein (**non-EncFtn**).

To understand how EncFtn and non-EncFtn relate to the well-characterised classical ferritins, seven classes of ferritins including FTN, BFR, DPS, mammalian ferritins (H-/M-/L-chains), the Er domain of Rr protein, EncFtn and non-EncFtn were aligned progressively based on two criteria: (1) the FOC is conserved in classical ferritins; (2) the C-terminal localisation sequences of EncFtn family have a common pattern. The multiple sequence alignment result was displayed in **Figure 7.2.1**. See **Section 2.10** for the method of the step-wise alignment and phylogenetic tree construction.

1.4.6.2 Evolutionary relationship of EncFtn and non-EncFtn

Based on the multiple sequence alignment (**Figure 7.2.1**) a phylogenetic tree (**Figure 1.4.9**) was built in MEGA7 (Kumar et al. 2016). It reveals an unambiguous classification of the different classes of the ferritin family, of which EncFtn class as represented by *Rhodospirillum rubrum* EncFtn (KEGG ID: Rru_A0973), stands for a unique ferritin clade. Based on the position of this group in the phylogenetic tree, it may be close to an ancestral form of ferritin. Interestingly, the non-EncFtn family may be closer to an ancestral ferritin than EncFtn as illustrated by the 'root' in **Figure 1.4.9**. The phylogenetic tree suggests EncFtn and non-EncFtn represent classes of ferritin that have diverged from the ancestral ferritin at an earlier stage than the classical ferritins; and the classical ferritin fold may have evolved from an ancestral form of ferritin that resembled the non-EncFtn or EncFtn through gene duplication and the introduction of a linker, ligating two monomers of EncFtn (Andrews 2010).

There is a notable outlier in the EncFtn family proteins. In *P. furiosus*, and a number of other archaeal species, a single gene (*PFC_05175*) encodes a polypeptide that is made of N-terminal EncFtn fused with C-terminal encapsulin in a single open reading frame; furthermore, the C-terminal location sequence is missing in this fused EncFtn and Enc (**Figure 7.2.1**). It is worth pointing out that only the EncFtn domain of PFC_05175 (PFC_01575short) has been used in the alignment to increase the alignment confidence. Therefore, this phylogeny of EncFtn does not show the relations of PFC_05175-like proteins with fused EncFtn and encapsulin to other EncFtn families. Given that the PFC_05175-like proteins are only found in archaeon, it is possible that they stand for a unique evolutionary clade.

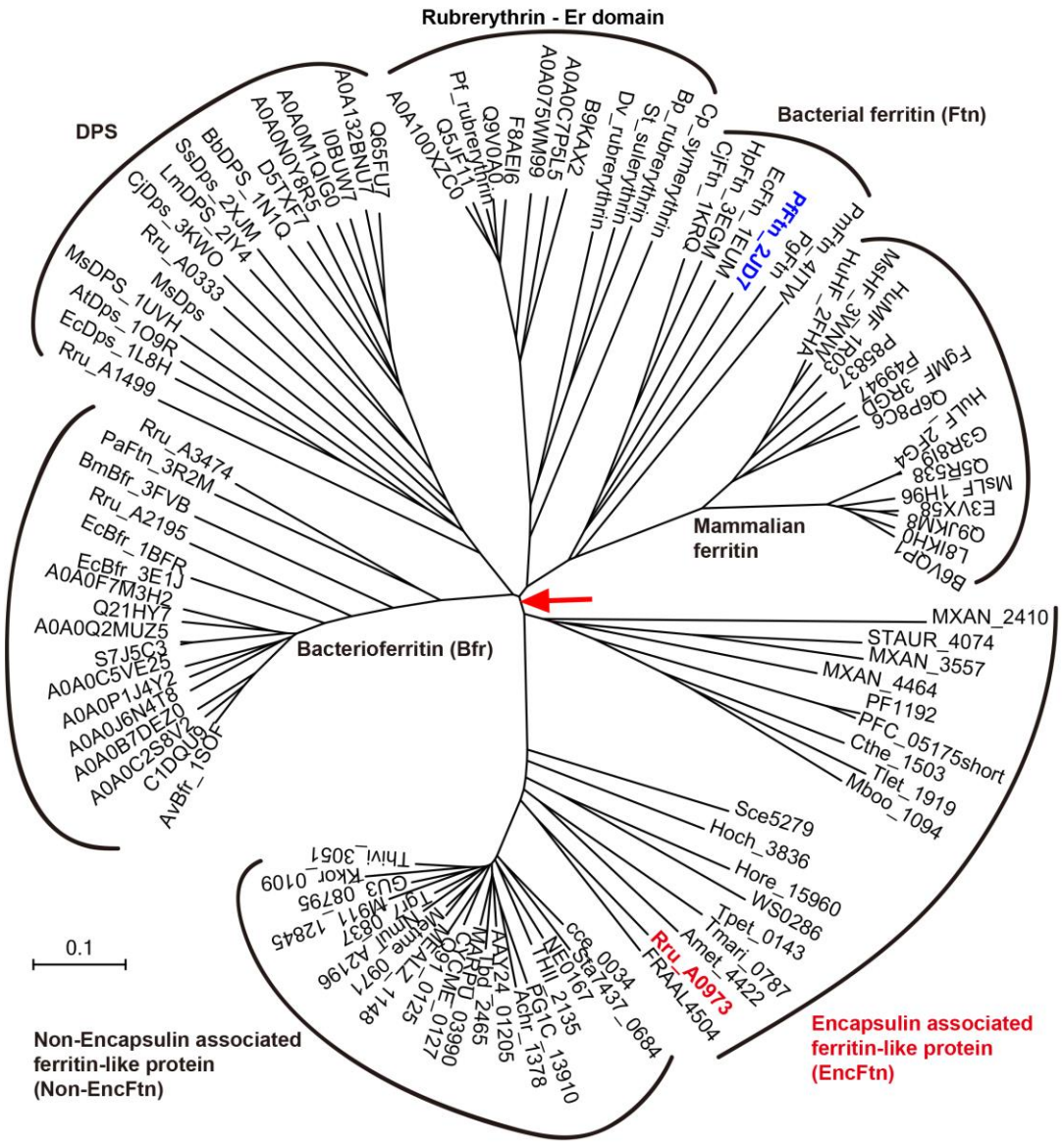


Figure 1.4.9 Phylogenetic tree of ferritin family proteins.

A phylogenetic tree of ferritin family proteins was built using the Neighbour-Joining method (Saitou & Nei 1987) based on a step-wise amino acid sequence alignment of the four-helical bundle portions of ferritin family proteins (**Figure 7.2.1**). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree; the likely root of the tree is indicated by a red arrow. The evolutionary distances were computed using the p-distance method (Nei & Kumar 2000) and are in the units of the number of amino acid differences per site. The rate of variation among sites was modelled with a gamma distribution (shape parameter = 2.5). The analysis involved 104 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 262 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016).

1.4.7 Biotechnology applications of ferritins

Ferritins have a long history of applications in biotechnology due to its biocompatibility (Charlton et al. 2016), size, self-assembly (Zhang & Orner 2011) and stable architecture, which is easy to engineer and modify for useful purposes (He & Marles-Wright 2015). Iron-loaded (recombinant) ferritins have been used as contrast agents for medical imaging, such as electron cryo-tomography (Wang et al. 2011) and magnetic resonance imaging (Jin et al. 2014), since the iron mineral core inside ferritin possesses superparamagnetic property that enhances the nuclear relaxation rate (Chasteen & Harrison 1999). Apart from delivering exogenous iron nanoparticles to biological systems, cellular contrast can also be induced *in vivo* by coexpressing ferritin and transferrin receptor, which has been shown to increase the targeting efficiency of inheritable MRI reporter to nonvascular cells and even brain tissue (Deans et al. 2006). As nanoparticles have a range of potential applications in nanotechnology, ferritins can also be used to produce metallic nanoparticles; application of these are reviewed in detail in a number of publications (He & Marles-Wright 2015; Zeth et al. 2016). For example, both *P. furiosus* bacterial ferritin (Kasyutich et al. 2010) and engineered human H-chain ferritin (Butts et al. 2008) have been employed to produce silver nanoparticles which have an antibacterial effect. Some metallic nanoparticles, such as Pt (Fan et al. 2011), Pd (Abe et al. 2008), Cu (Ceolín et al. 2008), Au (Fan et al. 2010) and TiO_2 (Klem et al. 2008) can be produced from horse spleen apoferritin or other mammalian ferritin sources and work as chemical catalysts. But it remains to be seen if the industrial production is possible and if they offer any material advantages over chemically produced nanoparticles. Ferritin contains three types of interfaces—the interior surface, the exterior surface and intersubunit interface, each of which can be genetically/chemically modified to enhance the functionality of the protein complex. The outer surface of *Helicobacter pylori* bacterial ferritin has been engineered to display haemagglutinin (HA), which can generate 10 times more antibodies with improved potency than the traditional influenza vaccines due to the adjuvant effect achieved by the localisation of the HA displayed on the ferritin (Kanekiyo et al. 2013). Surface-modified HuHF with an RGD peptide attached is applied as a drug (doxorubicin) carrier to target tumours as RGD peptide specifically binds a tumour angiogenesis marker—integrin $\alpha_v ß_3$(Zhen et al.

2013). The ability of phosphate incorporation into ferritin has been utilised to remove phosphate contamination in water by *P. furiosus* ferritin at an industrial-scale (Jacobs et al. 2010).

1.5 Encapsulin nanocompartment

1.5.1 Encapsulin structure and function

The first encapsulin protein was identified as a bacteriocin (linocin M18) from *Brevibacterium linens* (Valdés-Stauber & Scherer 1994). Its bacteriostatic effect was called into question because *B. linens* itself was used in the bacteriocin assay instead of pure encapsulin. Besides, the encapsulin homologues from *Mycobacterium tuberculosis* (Rosenkrands et al. 1998) and *Thermotoga maritima* (Hicks et al. 1998) have been tested regarding their bactericidal ability, and have both yielded negative results. The icosahedral encapsulin structures from the extremophilic archaeon *P. furiosus* (Pf-Enc) and hyperthermophilic bacterium *T. maritima* (Tm-Enc) were determined by X-ray crystallography from native expression (Akita et al. 2007; Sutter et al. 2008); while the encapsulin of the social bacterium *Myxococcus xanthus* (Mx-Enc) overexpressed in *E. coli* was structurally determined by Cryo-electron microscopy (Cryo-EM) (McHugh et al. 2014). Both the Pf-Enc and Mx-Enc consist of 180 subunits (12 pentamers plus 20 hexamers) and have $T = 3$ symmetry and 32–36 nm diameter, while the Tm-Enc possesses only 60 subunits formed by 12 pentamers with $T = 1$ symmetry and 24 nm diameter (**Figure 1.5.1**).

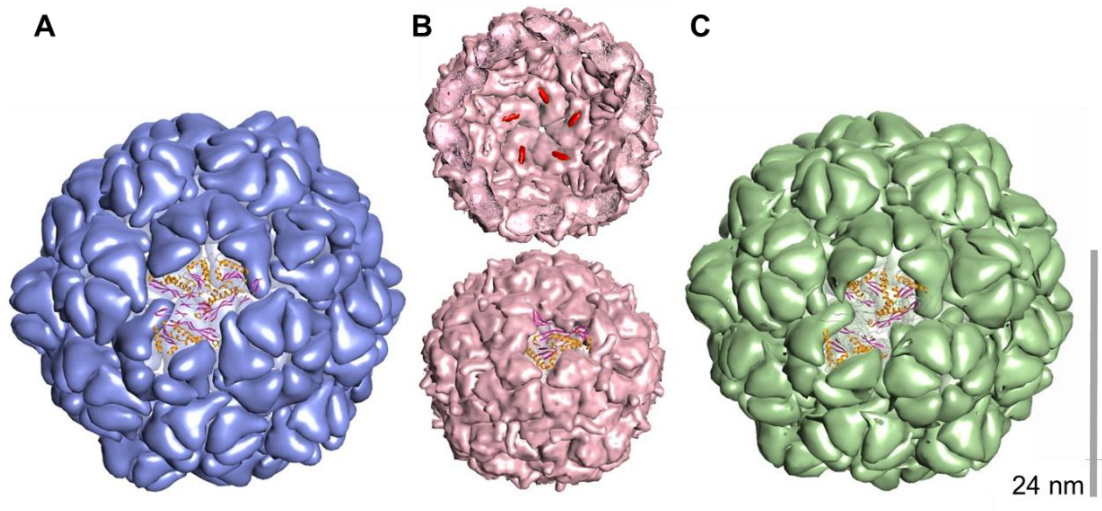


Figure 1.5.1 Protein structures of encapsulin nanocompartments.

Encapsulin nanocompartments are presented in surface mode with cartoons highlighting the crystallographic asymmetric unit. (A) 180-meric encapsulin from *P. furiosus* (PDB ID: 2E0Z), *T* = 3, diameter 32–36 nm. (B) 60-meric encapsulin from *T. maritima* (PDB ID: 3DKT), *T* = 1, diameter 24 nm. The top representation shows the cut-off view, in which the red surface shows the C-terminal localisation sequence of the native EncFtn enclosed within the encapsulin. (C) 180-meric encapsulin from *M. xanthus* (PDB ID: 4PT2), *T* = 3, diameter 32 nm.

Tm-Enc is suggested to house no more than 12 copies of the decamer EncFtn based on the observation of C-terminal localisation sequences of EncFtn targeting the inner surface of the capsid (Sutter et al. 2008). Mx-Enc shows the presence of about 30,000 iron ions and phosphate ions with a molar ratio of iron: phosphorus as 4:1 inside the cage, which is reminiscent of the iron mineral core inside ferritin cages (McHugh et al. 2014). Although there is no direct evidence of the enzymatic activities of encapsulins, it is reasonable to deduce the physiological role of the encapsulin nanocompartment based on the cargo proteins housed inside encapsulin shell (Giessen & Silver 2016).

1.5.2 Diversity of encapsulin nanocompartments

So far 909 encapsulins have been found in 15 bacterial and 2 archaeal phyla with various habitats and lifestyles through systematic genome analysis (Giessen & Silver 2016). The encapsulins house six types of cargo proteins: EncFtn (Sutter et al. 2008); dye-decolouring peroxidase (Dyp) (Rahmanpour & Bugg 2013); ferredoxin and hemerythrin (Giessen & Silver 2016); a ferritin homologue (BfrB), biophenyl dehydrogenase (BphB) as well as 7,8-dihydroneopterin aldolase (FolB) that is involved in folate biosynthesis (Contreras et al. 2014). These cargo proteins share the same feature with a characteristic C-terminal localisation sequence that can be recognised by certain types of encapsulins; in the case of ferredoxin the N-terminal sequence is targeting the encapsulin instead. It is worth mentioning that cargo proteins do not necessarily locate next to encapsulin in the genome. The number of cargo proteins encapsulated by one encapsulin nanocompartment ranges from 1 to 4, with different combinations of cargo species mentioned above (Giessen & Silver 2016). The presence of diversified encapsulin nanocompartments presents distinctive evolutionary advantages for microbes living in different habitats.

To understand the functions of EncFtn the work presented in this thesis will focus on the EncFtn-encapsulin nanocompartments (EncFtn-Enc). The three typical types of EncFtn-Enc in terms of genomic context will be analysed based on a KEGG taxonomy search (Ogata et al. 1999). Thus it is possible that other microbes not deposited in KEGG database have not been considered.

First of all, there are 34 species of bacteria with *EncFtn-Enc* genomic pattern over 3723 bacterial species across 10 classes according to KEGG taxonomy, as exemplified by Tm-Enc, possess an encapsulin protein with an EncFtn protein docked inside through its C-terminal sequence (**Figure 1.5.1 B**). The C-terminal sequences of the EncFtn family proteins share similar motifs (LT/GI/VGSLK) (**Figure 1.5.2**). In these bacterial genomes, the *encapsulin* gene locates downstream of the *EncFtn* gene with a short DNA sequence in between the two genes; and this linker ranges in lengths,

suggesting the possibility of an independent promoter for the encapsulin in some species.

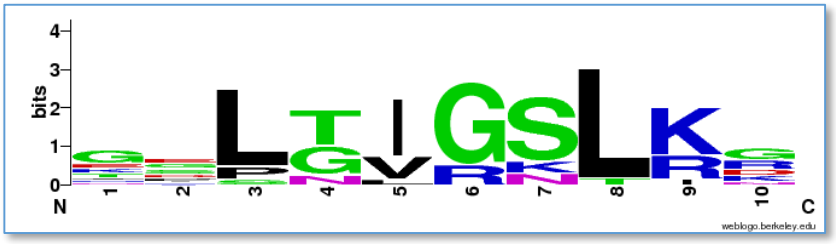


Figure 1.5.2 WebLogo of C-terminal localisation sequence of EncFtn family proteins.

EncFtn family proteins were aligned by Clustal Omega (Sievers et al. 2011; Li et al. 2015). A WebLogo of the aligned C-terminal localisation sequences was generated (Crooks et al. 2004), indicating that the 'LT/GI/VGSLK' sequence is well conserved in EncFtn family proteins.

Secondly, a single gene encoding a single protein with fused EncFtn and encapsulin is a shared feature for archaeal encapsulin nanocompartments, as illustrated in *P. furiosus* strain COM1 (PDB ID: 2E0Z) (Namba et al. 2005; Akita et al. 2007). Twenty-six archaeal species over 228 total archaea across three kingdoms have been found to possess this gene pattern based on the KEGG taxonomy (Ogata et al. 1999).

The third class of EncFtn-Enc, found in *M.* xanthus, where three different EncFtn proteins are found associated with the encapsulin protein (MXAN_3556) (McHugh et al. 2014). These three proteins all possess the C-terminal localisation sequence (MXAN_3557, MXAN_4464 and MXAN_2410), but only the *MXAN_3557* gene is locating next to the encapsulin gene. Among three EncFtn proteins only MXAN_2410 does not possess the conserved FOC residues (McHugh et al. 2014). The genomic location of the EncFtn proteins are not necessarily adjacent to the encapsulin gene in the genome. In the *Myxococcus fulvus* genome, genes *LILAB_30455* (*EncFtn*) and *LILAB_25765* (*Enc*) are widely separated; as are the ones found in the *Archangium gephyra* genome, *AA314_04443* (*EncFtn*) and *AA314_05059* (*Enc*). Some bacterial genomes also harbour more than one encapsulin nanocompartment locus; these may be induced in response to different environmental signals, or have different cellular functions. *Eubacterium limosum* contains *ELI_2650* (*EncFtn*)- *ELI_2651* (*Enc*) and *ELI_0383* (*EncFtn*)- *ELI_0384* (*Enc*). The two EncFtn proteins share 83% protein sequence identity, while the encapsulins share 71% protein sequence identity.

1.5.3 Applications of encapsulins

The encapsulin nanocompartment resembles a virus capsid in terms of its structural and mechanical properties. The self-assembly of encapsulin has been studied by native mass spectrometry as addition of dimers; and the encapsulation of native protein or heterologous protein with the same localisation sequence reduces the stability of encapsulin nanocompartments (Snijder et al. 2016). Heterologous cargo proteins such as fluorescent proteins have been co-expressed with *T. maritima* (Cassidy-Amstutz et al. 2016) or *Rhodococcus erythropolis* encapsulins (Tamura et al. 2014) by appending the characteristic C-terminal localisation sequences of EncFtn or Dyp, resulting in the encapsulation of these heterologous proteins. The *Rhodococcus jostii* encapsulin has been shown to disassemble at pH 3.0 and reassemble at pH 7.0, allowing the *in vitro* encapsulation of heterologous cargo with the localisation tag (Rahmanpour & Bugg 2013). A short peptide with a high affinity towards Fc region of immunoglobulin G (FcBP) has been successfully inserted into encapsulin sequence, allowing the display of this peptide on the encapsulin outer surface. These FcBP-Enc nanocompartments conjugated with fluorescent probe can be applied to visualise specific cancer cells (Moon et al. 2014). The encapsulin surface has also been used to display a viral haemagglutinin in order to produce antibodies with higher potency and broader influenza virus immunity than traditional vaccines (Kanekiyo et al. 2015). It is certainly worth exploring more biotechnological applications of encapsulins, due to their range of functions, stability and ease of modification and the analogies between encapsulin and classical ferritins.

1.5.4 Relationship between encapsulin nanocompartments and ferritins

In addition to the presence of genes encoding encapsulin and EncFtn, bacteria with these loci still possess genes encoding other ferritin family proteins, ranging from none (eg. *Wolinella succinogenes*, *Clostridium stercorarium*, *Eubacterium limosum* and *Halothermothrix orenii*) to 0–2 DPS, and 0–5 BFR/FTN based on a KEGG taxonomy search. It is not clear why microbes produce such a large complex when other ferritins are also present in their genomes. The function of encapsulin nanocompartments is likely to be related to the enclosed enzyme; that is protecting cells from toxic intermediates in microbes (Giessen & Silver 2016).

1.6 Aims

Encapsulin nanocompartments have been found in at least 1% of all bacteria and 10% of archaea from distinct ecological niches, with no clear link in terms of their lifestyles. Understanding the structure and function of encapsulin nanocompartments will aid in understanding the basic biochemistry and physiological role of encapsulins and inform future work on their application as platforms for biotechnology applications. In this thesis, an encapsulin nanocompartment from *Rhodospirillum rubrum* will be characterised structurally and functionally to understand the biochemistry of these intriguing protein complexes. *R. rubrum* is a versatile bacterium which can live aerobically by respiration in the dark or anaerobically via photosynthesis (Thore et al. 1969). The structures of EncFtn homologues from a halophilic bacterium *Haliangium ochraceum* (Fudou et al. 2002), and the extremophilic archaeon *P. furiosus*, will also be investigated to compare the EncFtn proteins from organisms from distinct habitats. The key research questions that will be addressed in this thesis are:

- What is the structure of EncFtn?
- How does EncFtn assemble?
- How does the EncFtn structure justify its function?
- What is the function of the encapsulin nanocompartment of *R. rubrum*?
- Do EncFtn proteins from different microbes possess different structures and functions?

Chapter 2 Materials and methods

All chemicals used in this work were purchased from Sigma-Aldrich unless otherwise stated. Plasmids pET-28a(+) and pACYCDuet1 used in the study were purchased from Novagen. Plasmid pET-28a-GG-RFP was constructed in the Marles-Wright Lab, which is based on modified pET-28a(+) with BsaI-RFP-BsaI replacing the original multiple cloning sites for compatibility with the CIDAR MoClo standard (Iverson et al. 2016). *R. rubrum* ATCC 11170 genomic DNA was purchased from DSMZ (Germany).

2.1 Cloning

Open reading frames encoding proteins from *R. rubrum* were amplified by PCR using *R. rubrum* genome as the template and KOD Hot Start DNA Polymerase (Novagen). The amplification of genes *Hoch-EncFtnS* (*Hoch_3836*$_{1\text{-}98}$) and *Pfc-EncFtnS* (*PFC_05175*$_{1\text{-}99}$) are based on gene templates codon optimised for *E. coli* expression and produced as double stranded gBlocks (IDT): *Hoch_3836* and *PFC_05175*, respectively. Primers used in this study are listed in **Table 2.1.1** and **Table 2.1.2**. PCR reactions were carried out according to the manufacturer's instructions. The resulting PCR product was visualised on a 1% agarose TAE gel and purified before digestion by appropriate restriction enzymes. The digested DNA was ligated with the similar digested plasmid before transforming *E. coli* TOP10 competent cells. The DNA insertion was confirmed by sequencing the plasmid extracted from transformed cells. All cloning except for Hoch-EncFtnS was conducted using traditional cloning method while in the case of Hoch-EncFtnS MoClo cloning with digestion and ligation in one pot by the Golden Gate protocol was applied.

2.1.1 Plasmid information

All constructs were made based on plasmids pACYCDuet-1, pET-28a(+) or pET-28a-GG-RFP (**Figure 2.1.1**). Constructs of encapsulin nanocompartment-related proteins were shown in **Table 2.1.1**. Constructs of *R. rubrum* EncFtn mutants were shown in **Table 2.1.2.**

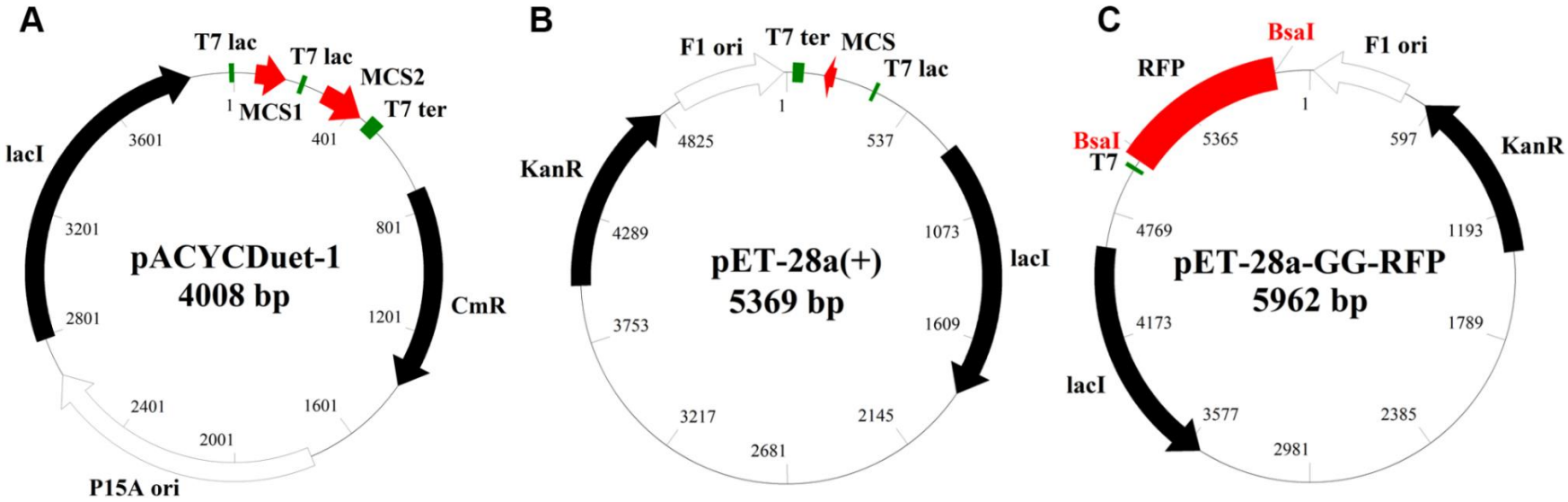


Figure 2.1.1 Plasmid maps.

Three maps of plasmids (A) pACYCDuet-1, (B) pET-28a(+) and (C) pET-28a-GG-RFP were generated by Bioedit software (Hall 1999). The red arrows in each plamid highlights the multiple cloning site for insertion of DNA. pACYCDuet-1 has chloramphenicol resistance casette, while both pET-28a(+) and pET-28a-GG-RFP harbour a kanamycin resistance casette.

2.1.2 Construct design

The short construct of Rru-EncFtn, Rru_A0973$_{1\text{-}96}$ with C-terminal hexahistidine tag (Rru-EncFtnSH) was designed based on a multiple sequence alignment of ferritin-like proteins (Error! Reference source not found.). The C-terminal region was predicted to b e unstructured in the structure of protein NE0167 from *N. europaea* (PDB ID: 3K6C) (Chang et al. n.d.). The C-terminal sequence of EncFtn was also suggested to be involved in docking into the inner surface of encapsulins in *T. maritima* (Sutter et al. 2008) and *M. xanthus* (McHugh et al. 2014).

Protein acronym	Construct	Forward primer	Restriction sites	Reverse primer	Restriction sites	Tag
Rru-EncFtnSH	pET-28a-Rru_A0973$_{1\text{-}96His6}$	GGCCCATG**GCGCAGTCGAG CAATAGCACC**	NcoI	AAGGAATTCGC**CGCGGT GATCGGCCCCTC**	EcoRI	C-terminal histidine*6
Rru-EncFtn	pACYCDuet-1-Rru_A0973	GGCCCATG**GCGCAGTCGAG CAATAGCACC**	NcoI	GGCGAATTC**TTACAACC GGGGTGGACGGG**	EcoRI	None
Rru-Enc	pACYCDuet-1-Rru_A0974	GTGTATTCCAT**ATGAACGA TCTGATGCGTGACTTG**	NdeI	TATGGTACC**TTATGCCT TGGTGGCGGC**	KpnI	None
Rru-EncFtn-Enc	pACYCDuet-1-Rru_A0973-Rru_A0974	GGCCCATG**GCGCAGTCGAG CAATAGCACC**	NcoI	GGCGAATTC**TTACAACC GGGGTGGACGGG**	EcoRI	None
		GTGTATTCCAT**ATGAACGA TCTGATGCGTGACTTG**	NdeI	TATGGTACC**TTATGCCT TGGTGGCGGC**	KpnI	None
Hoch-EncFtnS	pET-28a-GG-Hoch_3836$_{1\text{-}98}$	GGCTTGGTCTCAA**ATGGGC AGCAGCGAACAGCTG**	BsaI	GCGTAGGTCTCTAA**GCT TATTCTTCCAGTTCCAG AATCGGACGTTC**	BsaI	None
Pfc-EncFtnS	pET-28a-PFC_05175$_{1\text{-}99}$	GATCCATGGGC**CTGAGCAT CAACCCCACCCT**	NcoI	GGCGGATCCTTA**GGCCT CGATGCCGGTCAG**	BamHI	None

Table 2.1.1 Constructs of encapsulin nanocompartment related proteins.

Bases underscored in primers highlights the recognition site of endonuclease. Bases in bold are complementary to the DNA sequence of their target genes.

Rru-EncFtnSH or Rru-EncFtn-Enc based mutants				
Comment	**Template**	**Mutant name**	**Forward Primer**	**Reverse Primer**
Single mutant	pACYCDuet-1-Rru_A0973-Rru_A0974 or pET-28a-Rru_A0973$_{1\text{-}96His6}$	E32A	GTCGGTCATGGAGGCACTGGAAGCCGTCG	CGACGGCTTCCAGTGCCTCCATGACCGAC
		E62A	CATAATCGCGACGAGGCAAAAGAACACGC	GCGTGTTCTTTTGCCTCGTCGCGATTATG
		H65A	GACGAGGAAAAAGAAGCCGCGGCGATGACC	GGTCATCGCCGCGGCTTCTTTTTCCTCGTC
		E32D	GTCGGTCATGGAGGATCTGGAAGCCGTCG	CGACGGCTTCCAGATCCTCCATGACCGAC
		E62D	CATAATCGCGACGAGGATAAAGAACACGC	GCGTGTTCTTTATCCTCGTCGCGATTATG
		E31A	CGTGTCGGTCATGGCGGAACTGGAAGCCG	CGGCTTCCAGTTCCGCCATGACCGACACG
		E34A	CATGGAGGAACTGGCAGCCGTCGATTGGTAC	GTACCAATCGACGGCTGCCAGTTCCTCCATG
		W38G	CTGGAAGCCGTCGATGGGTACGACCAGCGCG	CGCGCTGGTCGTACCCATCGACGGCTTCCAG
Double mutants	pACYCDuet-1-Rru_A0973-Rru_A0974-E32A or pET-28a-Rru_A09731$_{\text{-}96His6}$-E32A	E32A+E62A	CATAATCGCGACGAGGCAAAAGAACACGC	GCGTGTTCTTTTGCCTCGTCGCGATTATG
	pACYCDuet-1-Rru_A0973-Rru_A0974-E62D or pET-28a-Rru_A09731$_{\text{-}96His6}$-E62D	E62D+E32D	GTCGGTCATGGAGGATCTGGAAGCCGTCG	CGACGGCTTCCAGATCCTCCATGACCGAC
	pACYCDuet-1-Rru_A0973-Rru_A0974-W38G or pET-28a-Rru_A0973$_{1\text{-}96His6}$-W38G	W38G+E31A	CGTGTCGGTCATGGCGGAACTGGAAGCCG	CGGCTTCCAGTTCCGCCATGACCGACACG
Triple mutants	pACYCDuet-1-Rru_A0973-Rru_A0974-E32A or pET-28a-0973sH-E32A	E32A+E62A+H65A	GACGAGGCAAAAGAAGCCGCGGCGATGACC	GGTCATCGCCGCGGCTTCTTTTGCCTCGTC
	pACYCDuet-1-Rru_A0973-Rru_A0974-W38G+E31A or pET-28a-Rru_A0973$_{1\text{-}96His6}$-W38G+E31A	W38G+E31A+E34A	CATGGCGGAACTGGCAGCCGTCGATGGGTAC	GTACCCATCGACGGCTGCCAGTTCCGCCATG

Table 2.1.2 Constructs of Rru-EncFtnSH or Rru-EncFtn-Enc based mutants.

Mutants of Rru-EncFtnSH and co-expressed Rru-EncFtn-Enc were designed based on plasmids pET-28a-Rru-EncFtnSH and pAYCDuet-1-Rru-EncFtn-Enc using site-directed mutagenesis kit, respectively. The primer information was listed in the table.

2.1.2.1 Traditional restriction enzyme cloning

Genes of interest were amplified by PCR from genomic or existing plasmid DNA. Briefly, each 50 µl PCR reaction contains: 10 ng plasmid or 100 ng bacterial genome as the template; 1.5 µl of 10 µM forward primer; 1.5 µl of 10 µM reverse primer; 45.5 µl PCR Master Mix (5 µl of 10 *x* buffer for KOD Hot Start DNA Polymerase (Novagen), 3 µl of 25 mM $MgSO_4$, 5 µl of 2 mM dNTPs and 32.5 µl ddH_2O); 1 µl KOD Hot Start DNA Polymerase (1.0 U/µl). To optimise PCR reactions additional 1 µl DMSO or additional 25 mM $MgSO_4$ could be supplemented. Thermocycling conditions include 95°C and 2 min for polymerase activation, 95°C and 20 s for denaturation, [lowest primer T_m]°C and 10 s for annealing, followed by 10–20 s extension at 70°C. The final step of PCR involves 70°C and 2 min. PCR product was stored at 4°C before resolving on 1% agarose gel. PCR product was either purified by recovering from gels or directly using QIAquick PCR Purification kit (Qiagen). Both DNA fragments and vector were digested at 37°C for 1 hr with appropriate restriction endonucleases (Fermentas FastDigest) by double digestion. Cut vector was purified by recovering from a 1% agarose gel. Digested DNA was purified directly using QIAquick PCR Purification kit. Ligation of cut DNA and plasmid was achieved by mixing at 10:1 to 3:1 molar ratio at room temperature for 1 hr with the T4 ligase and ligase buffer (Thermo Fisher Scientific). 5 µl ligation product was transformed into 50 µl competent TOP10 cells and left on ice for 20–30 min before heat shock at 42°C for 1 min. Cells were cooled on ice for 2 min before supplementing with 200 µl LB medium and incubated at 37°C and 900 rpm for 1 h. All cells were spread on LB with antibiotic plates. Negative control of transformation was also set up in parallel without plasmid transformed. Plates were incubated at 37°C overnight. Single colonies were incubated in 10 ml LB supplemented with appropriate antibiotic for plasmid propagation. Plasmids were purified from the cells using the Qiagen Miniprep kit based on the manufacturer instructions. Gene insertion was confirmed by Sanger sequencing of the resulting plasmid (Edinburgh Genomics).

2.1.2.2 CIDAR MoClo cloning

The CIDAR MoClo cloning method is based on the Golden Gate method which uses Type IIs restriction enzymes, which cut outside their recognition site, to allow the one-step digestion and ligation of DNA fragments, as well as multipart assembly (Iverson

et al. 2016). Hoch-EncFtnS was amplified via PCR with BsaI recognition sites and appropriate four base-pair overhangs appended to the 5' of the sense/anti-sense primers. The PCR product was purified as described above, followed by digestion and ligation with pET-28a-GG-RFP in one tube. Briefly, 10 µl system contains 1 µl of ~ 90 ng/µl Hoch-EncFtnS (~ 300 bp), 1 µl of ~50 ng/µl pET-28a-GG-RFP (~ 6 kbp), 1 µl of 10*x* T4 ligase buffer, 2 µl of 5 U/µl T4 ligase, 1 µl of 10 U/µl BsaI and 2 µl ddH_2O. The reaction was run in a PCR machine with 25 cycles of 37°C, 1.5 min and 16°C, 3 min. Further digestion was carried out at 50°C for 5 min to deactivate the DNA ligase, with enzyme deactivation at 80°C for 10 min. 5 µl of the resulting product was transformed into 50 µl competent *E. coli* TOP10 cells as described above. Colonies without red fluorescence from the RFP reporter were screened for positive insertion, which was subsequently confirmed by plasmid isolation and Sanger sequencing.

2.1.2.3 Site-directed mutagenesis

Site directed mutagenesis was carried out to introduce mutations to genes encoding *R. rubrum* encapsulin nanocompartment related proteins. Forward and reverse primers were designed to be complementary to each other with single or multiple mutations introduced. A 15–20 bp flanking sequence on both 5' and 3' of the mutation sites were required for primer design. 50 µl PCR system contains 45.5 µl PCR Master Mix, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, 70 ng plasmid as template (pET-28a-Rru-EncFtnSH or pACYCDuet-1-Rru-EncFtn-Rru-Enc) and 0.5 µl of 1 U/µl KOD Hot Start Polymerase. PCR reactions were run in steps including 95°C, 2 min, 8–15 cycles of 95°C, 20 s, 57°C, 30 s, 70°C and 4 min. 10 µl product was transformed into 100 µl competent TOP10 cells as described in **2.1.2.1**. Mutations were confirmed by Sanger sequencing.

2.2 Protein production and purification

A single colony of *E. coli* BL21(DE3) or *E. coli* Tuner(DE3) cells, transformed with protein expression plasmid, was transferred into 10 ml LB medium, or 100 ml M9 minimal medium (MM), supplemented with appropriate antibiotic, such as 50 ng/µl kanamycin (in H_2O) for pET-28a(+) or pET-28a-GG-RFP based constructs, or 34 ng/µl chloramphenicol for pACYCDuet-1 based constructs dissolved in ethanol. The culture was then incubated overnight at 37°C with 200 rpm shaking. 10 ml overnight

pre-culture was then inoculated into 1 litre of LB medium and incubated at 37°C with 200 rpm shaking. Recombinant protein expression was induced at OD_{600} = 0.6 by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the temperature was reduced to 18°C for overnight incubation. Cells were pelleted by centrifugation at 4000 x *g* for 20 min at 4°C, and resuspended in a ten-fold (volume per gram of cell pellet) of PBS to wash cells before a second centrifugation at 4000 x *g* for 20 min. Cells were resuspended in 10-times (v/w) of appropriate lysis buffer for the purification method used (**Table 2.2.1**) and lysed by sonication on ice, with ten cycles of 30-second burst of sonication at 10 µm amplitude and 30 seconds of cooling. Cell lysate was clarified by centrifugation at 20,000 x *g*, 4°C for 30 min, followed by filtration using a 0.45 µm syringe filter (Millipore). All the subsequent protein purification steps were performed on an ÄKTA workstation in The Edinburgh Protein Production Facility. Buffers used in the study are listed in **Table 2.2.1.**

Buffer name	Composition
HisA	50 mM Tris-HCl, 500 mM NaCl, 50 mM imidazole, pH 8.0
HisB	50 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 8.0
QA	50 mM Tris-HCl (pH 8.0)
QB	50 mM Tris-HCl (pH 8.0), 1 M NaCl
GF	50 mM Tris-HCl (pH 8.0), 150 mM NaCl
CD	10 mM sodium phosphate, 150 mM NaF, pH 7.94
HEPES	50 mM HEPES-NaOH, 150 mM NaCl, pH 8.0

Table 2.2.1 Buffer list.

Buffers containing Tris and NaCl were prepared using 1 M Tris-HCl (pH 8.0) and 5 M NaCl stock solutions.

2.2.1 Preparation of minimal medium

To prepare the defined M9 minimal medium the following recipe was used. 1 litre of 5x M9 salts containing 64 g $Na_2HPO_4{\cdot}7H_2O$ or 85.49 g $Na_2HPO_4{\cdot}12H_2O$, 15 g KH_2PO_4, 2.5 g NaCl and 5 g NH_4Cl, pH 7.4 was autoclaved before using. 40% (w/v) glucose or 40% (v/v) glycerol was filtered through 0.22 µm sterile syringe filter (Millipore) and stored at −20°C before using in the medium. 2 M $MgSO_4$ and 1 M $CaCl_2$ were separately prepared and filtered before using. 1 litre of M9 minimal medium(MM) contains 1x M9 salts, 0.4% glucose or glycerol, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and distilled water. Antibiotic was supplemented as required.

2.2.2 Preparation of selenomethionine-labelled proteins

Selenomethionine (SeMet)-labelling has been widely used in X-ray crystallography as a basic method of experimental phasing (Hendrickson et al. 1990). Here protein (Rru-EncFtnSH) was labelled with SeMet which allows the metal and Se analysis by inductive coupled plasma mass spectrometry (ICP-MS), which was conducted by Dr. Emma Tarrant and Dr. Kevin J. Waldron in the Institute for Cell and Molecular Biosciences of Newcastle University. Protein of interest was expressed in methionine auxotrophic *E. coli* strain B834(DE3) and grown in SeMet medium. SeMet incorporation rate was estimated by subjecting the protein to liquid chromatography-mass spectrometry (LC-MS), which was conducted by Dr. David Clarke in the School of Chemistry, University of Edinburgh.

2.2.2.1 Selenomethionine medium

SeMet medium contains 1x M9 salts, 0.4% (w/v) glucose, 19 L-amino acids (except methionine) at 40 μg/ml of each and vitamin mixtures as shown in **Table 2.2.2**, **Table 2.2.3** and **Table 2.2.4**. SeMet was freshly dissolved just before the supplementation to the medium. 1 M $MgSO_4$ and 10x M9 salts were autoclaved separately before using. Stock solutions of glucose and vitamin mixtures were filtered through 0.22 μm syringe filter (Millipore), respectively, before using in the medium.

SeMet medium	**Stock solution**	1 M MgSO4	10 x M9	40% (w/v) glucose	Amino Acids I	Amino Acids II	Vitamins	10 mg/ml SeMet	H_2O
	Volume	2 ml	100 ml	10 ml	10 ml	10 ml	1 ml	4 ml	to 1 litre

Table 2.2.2 SeMet medium.

SeMet is always freshly dissolved before using in the final medium.

L-Amino acids I	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Pro	Ser	Thr	Val	H_2O
	0.2 g																to 50 ml
L-Amino acids II	Phe	Trp	Tyr	H_2O													
	0.2 g			to 50 ml													

Table 2.2.3 L-Amino acids I and II.

L-Amino acids I and II are prepared separately and dissolved for at least 1 hour before using in the medium.

Vitamin mixtures (1 mg/ml, 1000 x)	Niacinamide (nicotinic acid amide)	Pyroxidine monohydrochloride	Riboflavin	Thiamine	H_2O
	10 mg				to 10 ml

Table 2.2.4 Vitamin mixtures.

Vitamin mixtures are transferred to aliquots of 1 ml and stored at −20°C.

2.2.2.2 Production of selenomethionine-labelled protein

On day 1, plasmid of interest was transformed into methionine auxotrophic *E. coli* B834(DE3) competent cells and grown on LB plate with certain antibiotic at 37°C overnight. On day 2, a single colony was inoculated into 10 ml LB with appropriate antibiotic added and incubated at 37°C and 250 rpm overnight. On day 3, 100 ml LB was inoculated with 1 ml overnight culture and incubated at 37°C and 250 rpm for 2–3 hrs until OD_{600} achieved 0.2–0.4. *E. coli* cells were pelleted by centrifugation at 4000 x *g* for 10–15 min and washed twice in SeMet medium (without SeMet and specific antibiotic) in order to remove the trace methionine from the LB medium. Then the pellet was resuspended in 10 ml SeMet medium and inoculated into 2 flasks of 500 ml SeMet medium (without SeMet and specific antibiotic). Appropriate antibiotics were supplemented to the flask before removing 10 ml culture to a 50 ml Falcon tube and incubating alongside as a non-Met control. Subsequently, SeMet was added to the flask and incubated at 37°C for 6 hrs when glucose was used as carbon source or 12 hrs when glycerol was used until the OD reached ~ 0.6. It is optional to leave the flask at 4°C for 1 hr prior to the induction by 1 mM IPTG at 18°C overnight (16–20 hrs). The cold shock before induction sometimes increases the soluble expression of proteins. On day 4, SeMet-grown cells were harvested by centrifugation as mentioned above. It is worth noting that all SeMet waste requires disposal as toxic waste according to toxic waste disposal rules in the university.

2.2.3 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was applied to analyse the proteins purified from each purification step. 15% or 18% polyacrylamide separating gel and 5% stacking gels were cast using the Mini-PROTEAN Tetra Cell (Bio-Rad) according to **Table 2.2.5**. 5 µl protein samples or cells were mixed with 5 µl 5x SDS loading buffer (250 mM Tris-HCl, pH 6.8, 10%(w/v)

SDS, 30% (v/v) glycerol, 5% (v/v) ß-mercaptoethanol and 0.02% (m/v) bromophenol blue) and heated at 100°C for 10 min. Samples were centrifuged at 15,870 x *g* for 1 min before loading the supernatant to each well of SDS-PAGE gel, which was then run in 1x SDS-PAGE running buffer. 10 litres of 10x SDS-PAGE running buffer contains 300 g Tris base, 1440 g glycine, 100 g SDS, pH 8.3.

4 gels with 1 mm thickness				
Stacking gel		Separating gel		
Gel percentage (%)	5	Gel percentage (%)	18	15
H_2O (ml)	4.08	H_2O (ml)	3.12	5.52
30% Polyacrylamide (ml)	1.02	30% Polyacrylamide (ml)	14.4	12
1 M Tris-HCl (pH 6.8) (ml)	0.75	1.5 M Tris-HCl (pH 8.8) (ml)	6	6
10% SDS(µl)	60	10% SDS(µl)	240	240
10% Ammonium persulfate (µl)	60	10% Ammonium persulfate (µl)	240	240
TEMED (µl)	6	TEMED (µl)	9.6	9.6
Total volume (ml)	6	Total volume (ml)	24	24

Table 2.2.5 SDS-PAGE protocol.

This protocol is used according to 4 polyacrylamide gels of 1 mm thickness. 24 ml of 15% or 18% separating gel is prepared firstly and added to the casting glasses, followed by covering with 0.5 ml isopropanol for each gel. The isopropanol is decanted after separating gel solidifying in 30 min. Subsequently, 6 ml of 5% stacking gel is mixed and added to the top of separating gel. A comb is inserted to cast wells. The stacking gel solidifies in about 1 hr before comb can be removed for sample loading.

2.2.4 Immobilized metal ion affinity chromatography purification

Recombinant proteins produced with hexa-histidine tags were purified by Immobilized Metal Ion Affinity Chromatography (IMAC). Clarified cell lysate was loaded onto a 5 ml HisTrap^{TM} FF column (GE Healthcare) pre-equilibrated in HisA buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl and 50 mM imidazole) at 1.5 ml/min. Unbound proteins were washed from the column with HisA buffer in 5 column volumes (CVs), or until the A_{280} reading stabilised, at 2.5 ml/min. His-tagged proteins were then eluted by 50% HisA buffer and 50% HisB buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl and 500 mM imidazole) at 1 ml/min in 4 CVs. The IMAC column was then washed with 100% HisB and regenerated by washing with H_2O (2 CVs), 0.5 M NaOH (5 CVs), H_2O (5 CVs), and stored in 20% ethanol (5 CVs). Protein fractions containing the protein of interest were identified by SDS-PAGE and pooled and concentrated to 2 ml prior to gel-filtration purification.

2.2.5 Anion exchange chromatography purification

Recombinant proteins produced without affinity tags were purified by anion exchange chromatography (AEC) using a 12 ml Q Sepharose™ Fast Flow ion exchange column (packed in the JMW laboratory). Cells were lysed and pelleted as mentioned above in **section 2.2**. Crude extract was loaded onto a Q Sepharose column pre-equilibrated with QA buffer (50 mM Tris-HCl, pH 8.0). Unbound samples were washed with 5 CVs of QA buffer. Bound proteins were eluted with a linear gradient of 0–100% QB buffer (50 mM Tris-HCl, pH 8.0, 1 M NaCl) over 20 CVs at 2.5 ml/min. Fractions containing the protein of interest were identified by 15% or 18% acrylamide SDS-PAGE. Fractions were pooled and concentrated before subsequent purification by size-exclusion chromatography.

2.2.6 Size-exclusion chromatography purification

2.2.6.1 Calibration of S200 size-exclusion gel filtration chromatography column

To calibrate the HiLoad 16/600 Superdex 200 (S200) Size-Exclusion Chromatography (SEC) column, gel filtration calibration kits (low molecular weight and high molecular weight) were used according to the GE Healthcare Life Science product information booklet. The elution volume (V_e) of each protein sample with known molecular weight was recorded. The void volume (V_o) of the S200 SEC column was determined using Blue Dextran 2000. Then the K_{av} values were calculated for each protein sample according to the equation: $K_{av} = (V_e - V_0) / (V_c - V_0)$. V_c = geometric column volume. The final calibration curve was plotted based on K_{av} *vs.* log molecular weight. The molecular weight of unknown proteins can be estimated according to the calibration curve.

2.2.6.2 SEC purification

Around 2 ml of protein was injected onto an S200 16/60 SEC column (GE Healthcare) pre-equilibrated in GF buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl) at a flow rate of 1 ml/min. All chromatography steps were performed at 8°C. Protein fractions of 1.8 ml were collected after the column void volume (40 ml) in 96-well deep-well blocks. Fractions were analysed by 15% or 18% SDS-PAGE and those containing the protein of interest were pooled for subsequent experiments. The S200 column was cleaned

using 0.5 M NaOH (1.5 CVs) and then ddH_2O (2.5 CVs). 1.2 CVs of 20% ethanol was used as a bacteriostatic agent for storage of the column.

2.2.7 Sucrose cushion/gradient ultracentrifugation purification

Coexpressed encapsulin associated ferritin and encapsulin (Rru-EncFtn-Enc), encapsulin protein (Rru-Enc) were purified based on a modified protocol used by M. Sutter (Sutter 2008). GF buffer was used to resuspend the cell pellet and produce sucrose solutions at 38% and 50% (w/v) sucrose. Solutions of 10%, 20%, 30% and 40% (w/v) sucrose were prepared by diluting 50% (w/v) sucrose in GF buffer. Crude extract was obtained by centrifuging lysed cells at 34,000 x *g* and 4ºC for 20 min and filtered through 0.45 μm syringe filter. To remove RNA and ribosomal contamination, the crude extract was supplemented with 50 μg/ml RNase A and rotated at 10 rpm at room temperature for 2 hrs, followed by centrifugation at 48,300 x *g* and 4°C for 20 min and filtering through 0.45 μm syringe filter.

2.2.7.1 Sucrose cushion production

Around 28 ml protein sample was loaded to the top of 10 ml of 38% sucrose cushion (**Figure 2.2.1 A/B**) in to a 38.5 ml Ultraclear tube (Beckman) and centrifuged at 100,000 x *g* (rotor SW28, 23600 rpm) and 4ºC for 21 hrs in a Beckman Coulter Ultracentrifuge Optima XPN-100. Maximum acceleration and '5' deceleration mode was used to keep the minimal disruption when pelleting. Subsequent to ultracentrifugation 5 ml fractions were removed from top to bottom into 15 ml Falcon tubes and labelled as 1–7 from top to bottom. After 35 ml of fractions were removed, three fractions of 1 ml were removed into 1.5 ml Eppendorf tubes. The pellet was resuspended in ~9 ml GF buffer on a vortex mixer. The insoluble portion of the pellet was removed by centrifugation at 10,000 x *g* and 4ºC for 15 min in Beckman Avanti J-25 centrifuge with the rotor JA-25.50. Samples from each fraction were analysed by 15% or 18% SDS-PAGE and fractions containing the protein of interest were pooled for subsequent purification steps.

2.2.7.2 Sucrose gradient centrifugation

Sucrose gradient was set up in a 38.5 ml Ultraclear tube (Beckman) according to **Figure 2.2.1 A/C**. Around 13 ml of filtered supernatant from sucrose cushion ultracentrifugation was added gently to the top of the 10%–50% sucrose gradient.

Samples were split into multiple tubes for better separation of the protein of interest from contaminants. Protein samples on the sucrose gradient were centrifuged at 100,000 x *g* and 4°C for 17 hrs. Three 5 ml fractions were removed from the top of the tube; the remaining liquid was removed as 0.75 ml fractions and placed in 1.5 ml Eppendorf tubes. The protein content of each fraction was quantified by the measurement of absorbance at 280 nm (A_{280}) and protein purity was estimated based on the protein/DNA ratios (A_{260}/A_{280}) via NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific).

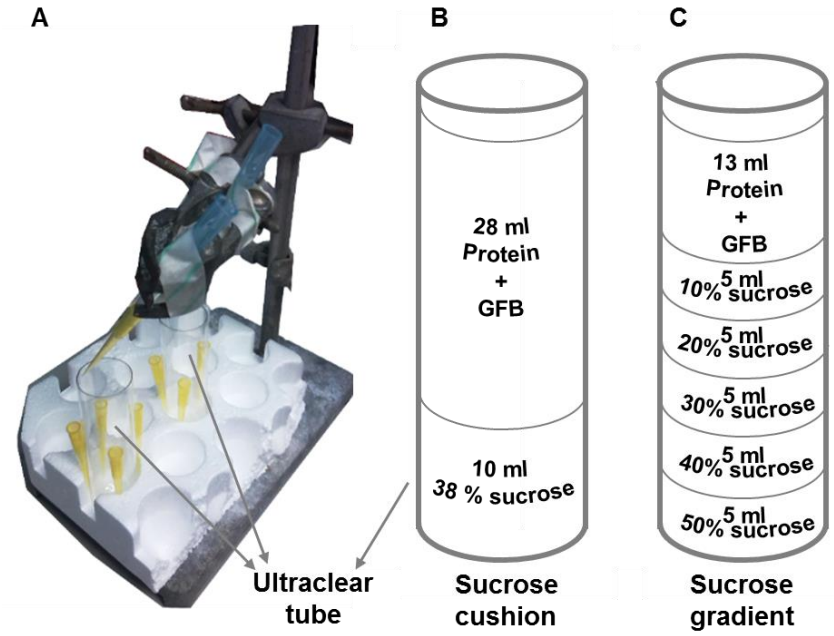


Figure 2.2.1 Sucrose cushion/gradient setup.

(A) Two 38.5 ml Ultraclear tubes were placed on a foam base with tips set up to allow the sucrose cushion (38%) or gradient (10%–50%) to be gently settled. (B) Sucrose cushion was set up in an Ultraclear tube with 10 ml of 38% sucrose in the bottom and 28 ml protein in GF buffer on the top. (C) Sucrose gradient was set up as five gradients of 5 ml sucrose ranging from 50% to 10% with 13 ml protein in GF buffer topping up.

2.3 Transmission electron microscopy

Negative stain transmission electron microscopy (TEM) imaging was performed on purified encapsulin, EncFtn, and co-purified encapsulin/EncFtn. Purified protein at 0.1 mg/ml concentration was spotted on glow-discharged 300 mesh carbon-coated copper grids and excess liquid wicked off with Whatman filter paper (grade 1, diameter 24.0 cm). The grids were washed with 10 μl distilled water and blotted with filter paper three times before staining with 2 % uranyl acetate, followed by blotting and air-drying.

Grids were imaged using a Philips CM120 transmission electron microscope and images were collected with a Gatan CCD camera. The microscope was later updated to JEOL JEM-1400 operating at 80 kV with a magnification up to 2,000,000X. The camera was updated to GATAN One View. Images were analysed using ImageJ (NIH) and size-distribution histograms were plotted using Prism6 (GraphPad software). Here uranyl acetate was used because the heave ions readily interacted with the electron beam and produced image contrast. When uranyl acetate was added to protein sample, the stain would surround the protein but was excluded from the protein volume. The electron beam can pass the protein sample easier than the surrounding space, leading to the observation of a dark background and the morphology of a white protein (Barreto-Vieira & Barth 2015).

2.4 Peptide mass fingerprinting

Proteins can be identified by peptide mass fingerprinting which involves three steps: protein separation by SDS-PAGE, trypsin digestion, and peptide identification by the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Thiede et al. 2005). Firstly, protein was separated on SDS-PAGE gel as single bands, which were then cut into 2 mm x 1 mm pieces. Prior to trypsin digestion, protein bands were washed in 300 µl of 200 mM NH_4HCO_3 (ABC) and 50% acetonitrile (ACN) at room temperature for 30 min; this was repeated twice. The protein band was then centrifuged at 15000 x *g* and the supernatant was discarded. The protein band was then incubated in 300 µl 20 mM DTT, 200 mM ABC, 50% ACN at room temp for 1 hour, followed by instant wash in 300 µl 200 mM ABC, 50% ACN for three times. Cysteines were alkylated by adding fresh 100 µl 50 mM iodoacetamide (IAA), 200 mM ABC, 50% ACN at room temperature in dark for 20 min. Protein band was subsequently washed in 500 µl 20 mM ABC, 50% ACN for 3 times. The protein band was then centrifuged at 15,870 x *g* for 2 min and remove the supernatant. Subsequently, the protein band was covered with ACN for at least 5 min until the gel turned white. Gel pieces were dried with tube cap open after decanting ACN. Trypsin stock was prepared by adding 50 µl of 50 mM ABC to a new vial of trypsin powder (Sigma-Aldrich) at 4°C, splitting into small aliquots and stored at −80°C. 30 µl trypsin working solution including 29 µl of 50 mM ABC and 1 µl trypsin stock solution was added to the gel-pieces on ice until the gel swells. A trypsin blank without protein

bands was set up alongside to provide an internal calibration peak for MALDI-MS. All tubes with trypsin digestion were sealed with Parafilm M and incubated at 32°C for 16–24 hrs.

On the second day, protein samples were centrifuged and sonicated for 10 min. 0.5 µl digested peptide was spotted on a MALDI gold plate with 0.5 µl α-cyano-4-hydroxycinnamic acid matrix (Sigma-Aldrich). When samples were analysed in MALDI-MS, peptide mass peaks were detected with baseline correction, noise filter, peak deisotoping as well as mass calibration against trypsin autolysis peak. Peaks were then searched on Mascot Server (Matrix Science) against 'Swiss-Prot' or 'NCBInr' databases for significant hits. Parameters were set up as following: 100 ppm as peptide tolerance, carbamidomethyl as fixed modifications and oxidation on methionine as variable modifications.

It is worth mentioning two points: (1) gloves should be worn to avoid the contamination; (2) small proteins might get an insignificant searching results, of which a comparison with virtual trypsin digestion result would be helpful.

2.5 Protein crystallisation, X-ray data collection and analysis

2.5.1 Protein crystallisation screening and optimisation

Protein purified through IMAC or AEC and then SEC was concentrated to around 10 mg/ml. Protein concentration was estimated based on calculated extinction coefficients at 280 nm (Simonian & Smith 2001). Proteins were screened for crystallisation against the following commercial crystallisation kits: PEG/ION screen from Hampton Research; JCSG-plus, Structure Screen, and MIDAS from Molecular Dimensions.

An Art Robbins Gryphon nanolitre pipetting robot was used to dispense 65 µl of each crystallisation screen from a master screen kit into a MRC 96-well crystallisation plate. 0.1 µl reservoir solution was pipetted into each crystallisation well, followed by 0.2 and 0.1 µl protein dispensed into the top and bottom crystallisation wells by the nanodispenser. The sitting-drop vapour-diffusion crystallisation plate was sealed by X-Seal Manual Adhesive Sealer XCS-384 (FluidX) using transparent and self-adhesive polyolefin film. Crystallisation plates were incubated at 18°C. Crystallisation

conditions were then optimised in Linbro 24-well plates based on protein crystal hits found in 96-well plates.

In optimisation trays, crystallisation drops were set up using the hanging drop vapour diffusion method using 24 well Linbro plates. Well solution was set up at 1 ml final volume. Glass coverslips were set up with the ratio of protein to well solution as 1 µl: 1 µl, 1 µl: 2 µl and 2 µl: 1 µl and wells were sealed with vacuum grease (Corning). Plates were stored at 292 K for one week to two months.

2.5.2 Crystal soaking and mounting

Crystals were harvested from 96 and 24 well plates using a LithoLoop (Molecular Dimensions Limited); transferred briefly to a cryoprotection solution containing well solution supplemented with 20 % (v/v) PEG 200; and subsequently flash cooled in liquid nitrogen. For iron (II) soaking, crystals were transferred into a cryoprotectant with 1 mM $FeSO_4$ or $Fe(NH_4)_2(SO_4)_2$ dissolved in 0.1 % (v/v) HCl for different time intervals and back soaked in a cryoprotectant without iron (II) prior to flash cooling in liquid nitrogen.

2.5.3 X-ray diffraction data collection and processing

All crystallographic datasets were collected on the macromolecular crystallography beamlines at Diamond Light Source (Didcot, UK) at 100 K using Pilatus 6M detectors. An X-ray fluorescence spectrum (MCA) was performed to indicate the presence of metal in the protein crystal. A fluorescence edge scan was carried out to precisely determine the peak and inflection points of the absorption edge of any anomalously scattering metals present in the crystal. Most diffraction data were collected at 10–100 eV above the theoretical Fe-*K* edge or experimentally determined E(peak) to maximise the anomalous peaks of iron, leading to a recordable resolution of 2 Å due to the limitation of crystal-to-detector distance. The triple mutant E32A-E62A-H65A of Rru-EncFtnSH was collected at 1.5 Å in order to increase the resolution limit when no iron was bound in this mutant.

Diffraction data were integrated and scaled using XDS (Kabsch 2010); the crystallographic point group and space group were determined by Pointless and symmetry related reflections were merged with Aimless (Evans 2011). The resolution

cut-off used for structure determination and refinement was determined based on the completeness (>95%), I/sigI (>1.5) and $CC_{1/2}$ (>0.5) criterion of Karplus and Diederichs (Karplus & Diederichs 2012).

2.5.4 Crystallographic model building and refinement

The structure of Rru-EncFtnSH was determined by molecular replacement using protein NE0167 (PDB ID: 3K6C) as the search model, which had been modified to match the sequence of the target protein using Chainsaw (Stein 2008). A single solution comprising three decamers in the asymmetric unit was found by molecular replacement using Phaser MR (McCoy et al. 2007). The initial model was rebuilt using Phenix.autobuild (Adams et al. 2010) followed by cycles of refinement with Phenix.refine (Afonine et al. 2012), with manual rebuilding and model inspection in Coot (Emsley et al. 2010). Each refinement cycle in Phenix.refine included ten macrocycles of bulk solvent correction, anisotropic scaling, coordinate refinement, individual B-factors and occupancy refinement. The refinement target weights included scale factor for X-ray/stereochemistry weight (wxc_scale) as 0.5; for X-ray/ADP weight (wxu_scale) as 1.0; stereochemistry weight scale (wc) as 1.0; ADP weight scale (wu) as 1.0. The refinement program has been repeated several cycles until the R_{work}, R_{free} and stereochemical restraints have achieved convergence. In the final stage of the refinement, water molecules and hydrogens were added to the protein model automatically in Phenix and further inspected in Coot manually. The final model was also refined with torsional NCS restraints (2.5° for restraint sigma and 15° for restraint limit), as well as anomalous group refinement if anomalous peaks have been detected.

Refmac5 had been used in the early stage of refinement (Pannu et al. 1998; Steiner et al. 2003) before Phenix.refine was used. As a fast refinement tool, refmac5 was able to produce a good starting model for further refinement with Phenix.refine. However, as two refinement programs have many differences such as weight factors, continually switching between refmac5 and phenix.refine would lead to unsuccessful refinement.

The structures of Rru-EncFtnSH mutants and homologues were determined by molecular replacement using a modified monomeric or decameric Rru-EncFtnSH structure as the template with refinement and model building as described above.

2.5.5 Crystallographic model analysis

The crystallographic model was validated using MolProbity (Chen et al. 2010). Structural superimpositions were calculated using the secondary-structure matching (SSM) (Krissinel & Henrick 2004) algorithm implemented in Coot (Emsley et al. 2010). Crystallographic figures were generated with PyMOL (Schrodinger 2015). Protein interface interactions were calculated using the PDBePISA server (Krissinel & Henrick 2007). The metals modelled in proteins have been validated by the CheckMyMetal server (Zheng et al. 2014). See appendices for the CheckMyMetal reports of all EncFtn proteins listed in this thesis.

2.6 Biochemical analyses and assays

2.6.1 Ferroxidase assay

Ferrous iron stock was made by dissolving 0.1 M $Fe(NH_4)_2(SO_4)_2$ in 0.1% (v/v) HCl in anaerobic conditions. Diluted Fe(II) solutions were also prepared anaerobically in 0.1% (v/v) HCl. Protein solutions were prepared by diluting purified protein samples in GF buffer (50 mM Tris-HCl, pH 8.0 and 150 mM NaCl) anaerobically. Ferroxidase activity was initiated by adding 450 µl protein to 50 µl of acidic $Fe(NH_4)_2(SO_4)_2$ at the final concentration of 100 µM and 20 µM in the air, respectively. The ferroxidase activity was measured by monitoring the Fe(III) formation which gives rise to the change of the absorbance at 315 nm (Bonomi et al. 1996). Absorbance at 315 nm was recorded every second over 1800 seconds using a quartz cuvette in a JASCO V-730 UV/VIS spectrophotometer (JASCO Inc.). In recombinantly coexpressed nanocompartments the ratio of EncFtn to Enc was assumed as 2 to 1, assuming each of the twelve pentameric vertices of the icosahedral encapsulin were occupied with decameric EncFtn. The data are presented as the mean of three technical replicates with error bars indicating one standard deviation from the mean. Proteins used here were from a single preparation.

2.6.2 Peroxidase assay

The peroxidase activity of Rru-EncFtnSH was determined by measuring the oxidation of a chromogenic substrate *ortho*-phenylenediamine (OPD) by H_2O_2 (Pesek et al. 2011). Rru-EncFtnSH decamer and monomeric fractions purified from minimal media were both used in the assay. OPD was prepared as a 92.5 mM stock solution in 50 mM

Tris-HCl (pH 8.0). Different concentrations of OPD were prepared by diluting the stock solution in 50 mM Tris-HCl (pH 8.0). 100 µl of each diluted OPD was added to a 96-well plate in 3 replicate wells. 1 µl of 32 µM protein was supplemented into each well to a final concentration of 160 nM, followed by the addition of 2 µl of 30% H_2O_2. After 15 minutes shaking in the dark, the reaction was stopped by adding 100 µl of 0.5 M H_2SO_4. The peroxidase activity was measured by monitoring the absorbance at 490 nm in the SpectraMax M5 Microplate Reader (Molecular Devices Corporation) (Pesek et al. 2011).

2.7 Iron content analysis of protein samples

2.7.1 Inductively Coupled Plasma Mass Spectrometry (ICP-MS) analysis

ICP-MS has been wildly used in determining isotropic composition in a semi-quantitative or quantitative manner. ICP-MS can detect most elements in the periodic table with detection limits below part per trillion. Protein samples were diluted 50-fold into a solution of 2.5 % HNO_3 (Suprapur, Merck) containing 20 µg/l Pt as internal standard. Matrix-matched elemental standards (containing analyte metal concentrations 0–1000 µg/l) were prepared by serial dilution from individual metal standard stocks (VWR) with identical solution compositions, including the internal standard. All standards and samples were analysed by ICP-MS using a Thermo X-series instrument operating in collision cell mode (using 3.0 ml min^{-1} flow of 8% H_2 in He as the collision gas). Isotopes ^{44}Ca, ^{56}Fe, ^{66}Zn, ^{78}Se, and ^{195}Pt were monitored using the peak-jump method (100 sweeps, 25–30 ms dwell time on 5 channels per isotope, separated by 0.02 atomic mass units) in triplicate. This experiments have been carried out by Dr. Emma Tarrant and Dr. Kevin J. Waldron in the Institute for Cell and Molecular Biosciences, Newcastle University.

2.7.2 Ferrozine assay

Protein	Final protein concentration per 1mer (uM)	100ul=5ul Fe(II) + protein + GFB								Label: 50
		+ 5 µl * 0.1% (v/v) HCl	Fe(II) loading (10-40 were numbers to label all tubes)							Control: 2.5mM Fe(II)
2 µl of **73sH**-SeMet+Fe in HEPES + 93ul GFB	8.46	Label: 10	Label	11	12	13	14	15		5ul * 40mM Fe(II) + 95ul GFB, desalt
			Fe stock (mM)	0.8	1.7	2.9	4.5	6.0		
			Final Fe loading (µM)	39.9	84.0	147.0	224.0	301.0		
			Fe/73sH-10mer (per 1mer)	5	10	17	26	36		
6 µl of **0974**-SeMet in GFB +89ul GFB	8.50	Label: 20	Label	21	22	23	24	25	26	27
			Fe stock (mM)	4.5	6.0	9.0	14.0	20.0	30.0	40.0
			Final Fe loading (µM)	224.0	301.0	450.0	700.0	1000.0	1500.0	2000.0
			Fe/0974 (per 1mer)	26	35	53	82	118	176	235
13 µl of **7374**-SeMet in GFB +82ul GFB	8.70	Label: 30	Label	31	32	33	34	35	36	37
			Fe stock (mM)	4.5	6.0	9.0	14.0	20.0	30.0	40.0
			Final Fe loading (µM)	224.0	301.0	450.0	700.0	1000.0	1500.0	2000.0
			Fe/7374 (per 1mer)	26	35	52	80	115	172	230
12 µl of **apo-HsFtn** in GFB +83 ul GFB	8.50	Label: 40	Label	41	42	43	44	45	46	
			Fe stock (mM)	0.0	0.9	4.3	12.8	25.5	34.0	
			Final Fe loading (µM)	0.0	42.5	212.5	637.5	1275.0	1700.0	
			Fe/apo-HsFtn (per 1mer)	0	5	25	75	150	200	

Table 2.7.1 Fe(II) loading experiment.

8.5 µM proteins including Rru-EncFtnSH decamer fractions (73sH), Rru-Enc (0974), coexpressed Rru-EncFtn and Rru-Enc (7374) and horse spleen apoferritin (apo-HsFtn) were mixed with $Fe(NH_4)_2(SO_4)_2$ in 0.1% (v/v) HCl at various concentrations at room temperature for 3 hrs, respectively. This system was topped up with GF buffer to 100 µl. Protein with 0.1% (v/v) HCl (label: 10, 20, 30, 40) and Fe(II) with GF buffer (label: 50) were both set up as controls. Protein samples were centrifuged, of which the supernatant was desalted to remove the free Fe in solutions. Fe control was also desalted to determine the desalting efficiency which turns out to be over 95%.

In order to determine the maximum iron loading capacity of different EncFtn and Enc samples an iron loading experiment was carried out according to **Table 2.7.1**. For the analysis of encapsulin compartments, 8.5 μM EncFtn-Enc and 8.7 μM Encapsulin were diluted in GF buffer and both were mixed with various concentrations of $Fe(NH_4)_2(SO_4)_2$ (prepared in 0.1% (v/v) HCl). To assay the iron loading capacity of EncFtn, 8.5 μM of decameric Rru-EncFtnSH was diluted in GF buffer and mixed with various concentrations of $Fe(NH_4)_2(SO_4)_2$ (prepared in 0.1% (v/v) HCl) as before. Protein and iron mixtures were incubated at room temperature for 3 hrs. Precipitated protein and iron oxides in the samples were removed by centrifugation at 15,871 x *g* for 10 min. Free iron ions in solution were removed through buffer exchange using Zeba™ spin desalting columns (7 kDa cut-off, Thermo Fisher Scientific). The protein concentration of each sample was determined using the Pierce Microplate BCA assay kit (Thermo Fisher Scientific). A protein standard curve was plotted according to the manufacturer instruction. The iron content in the samples was determined using a modified ferrozine assay (Riemer et al. 2004). In this assay 100 μl of protein sample was mixed with 100 μl mixture of equal volume of 1.4 M HCl and 4.5% (w/v) $KMnO_4$ and incubated at 60°C for 2 hrs. Prior to this incubation 20 μl of the iron-detection reagent (6.5 mM ferrozine, 6.5 mM neocuproine, 2.5 M ammonium acetate, and 1 M ascorbic acid dissolved in H_2O) was added to the cooled tube. After a 30 min incubation, 200 μl of the solution was transferred to a 96-well plate and the absorbance at 562 nm was measured on a plate reader Spectramax M5 (Molecular Devices). A standard curve for iron concentration was plotted using various concentrations of $FeCl_3$ (in 0.1% (w/v) HCl) diluted in the gel-filtration buffer (curve available in Figure 5.6.1). Both the ferrozine and BCA microplate assays were performed using three technical repeats. Samples analysed by ICP-MS were prepared in the same way by mixing protein and ferrous ions and desalting. The proteins used in the iron loading experiment came from a single preparation. Horse spleen apo-ferritin was incubated with Fe(II) according to **Table 2.7.1** as a positive control group. 2.5 mM Fe(II) was incubated with GF buffer alongside to detect the desalting efficiency of the Zeba™ spin desalting column.

2.8 Protein assembly analysis

2.8.1 SEC-MALS

Size-exclusion chromatography (ÅKTA-Micro; GE Healthcare) coupled with multi angle light scattering was used to determine the molecular mass of gel-filtration fractions of Rru-EncFtnSH in solution. Firstly, sample components were separated by SEC based on the hydrodynamic size, followed by running through UV detector, multiple light scattering detectors at fixed angles (Viscotek SEC-MALS 20) and refractive index detector (Viscotek VE3580, Malvern Instruments). The major advantage of gaining light scattering data as a function of scattering angle is the radius of gyration (Rg) can be calculated to give the size of the molecules. In the case of SEC, if the scattering molecules are relatively big and the light scattering varies with angle, which makes the Rg measurement important for the precise determination of the molar mass of the molecules when the anisotropy comes into play. The molar mass of the polypeptide portion of the protein sample (MM_{pp}) can be calculated as follows (Folta-Stogniew 2006):

$$MM_{pp} = K * (LS)(UV) / [A * (RI)^2]$$

Where K is the calibration constant of the SEC-MALS system as determined from the calibration curve generated using the same protein buffer and BSA protein standards; LS is the response from light scattering detector; UV equals to A_{280}; RI is the response from the refractive index detector; A is the extinction coefficient of the polypeptide and expressed as $A^{0.1\%}_{280}$ (mg/ml).

Here the protein concentration was determined by measurement of absorbance at 280 nm using a calculated extinction coefficient ($\varepsilon^{0.1\%}$) of 1.462 mg^{-1} ml cm^{-1} for Rru-EncFtnSH. 100 µl of 1.428 mg/ml decameric fraction and 4.027 mg/ml monomeric fraction were run individually on a Superdex 200 10/300 GL size exclusion column pre-equilibrated in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl at 22 ˚C with a flow rate of 0.5 ml/min. Light scattering, refractive index (RI) and A_{280nm} were analysed by a homo-polymer model (OmniSEC software, v 5.1; Malvern Instruments) using the following parameters: the extinction coefficient (dA/dc) at 280 nm was 1.46 AU ml mg^{-1} and specific refractive index increments (dn/dc) was 0.185 ml g^{-1}.

2.8.2 Mass spectrometry

For native Mass Spectrometry or nano-electrospray ionisation mass spectrometry (nESI-MS) analysis, all protein samples were buffer exchanged into 100 mM ammonium acetate (pH 8.0; adjusted with dropwise addition of 1% ammonia solution) using Micro Biospin® Chromatography Columns (Bio-Rad, UK) prior to analysis and the resulting protein samples were analysed at a final concentration of ~ 50 μM (monomer concentration). In order to obtain iron-bound Rru-EncFtnSH, 100 μM or 300 μM of freshly prepared $FeCl_2$ was added to apo-Rru-EncFtnSH (monomer fractions) immediately prior to buffer exchange into 100 mM ammonium acetate (pH 8.0). Samples were analysed on a quadrupole ion-mobility time of flight instrument (Synapt G2, Waters Corp., Manchester, UK), equipped with a nanomate nanoelectrospray infusion robot (Advion Biosciences). Instrument parameters were carefully tuned to preserve non-covalent protein complexes. After optimisation, typical parameters were: nanoelectrospray voltage 1.54 kV; sample cone 50 V; extractor cone 0 V; trap collision voltage 4 V; source temperature 80°C; and source backing pressure 5.5 mbar. For improved mass resolution the sample cone was raised to 155 V. Ion mobility mass spectrometry (IM-MS) was performed using the travelling-wave mobility cell in the Synapt G2, employing nitrogen as the drift gas. Typically, the ion mobility separation (IMS) wave velocity was set to 300 m/s; wave height to 15 V; and the IMS pressure was 1.8 mbar. For collision cross section determination, IM-MS data was calibrated using denatured equine myoglobin and data was processed using Driftscope v2.5 and MassLynx v4.1 (Waters Corp., UK). Theoretical collision cross sections (CCS) were calculated from pdb files using IMPACT software v. 0.9.1 (Marklund et al. 2015). In order to obtain information on the topology of the EncFtn assembly, gas-phase dissociation of the Fe-associated EncFtn complex was achieved by increasing the sample cone and/or trap collision voltage prior to MS analysis. All nESI-MS and IM-MS were performed by Dr. David Clarke and Sam Huges in the School of Chemistry, The University of Edinburgh.

2.8.3 Native PAGE analysis of proteins

Native polyacrylamide gel electrophoresis (PAGE) or nondenaturing PAGE was applied to analyse the protein oligomeric states in native state. A 12% native polyacrylamide gel was cast using 5.6 ml of 30% polyacrylamide solution, 3.5 ml of

1 M Tris-HCl (pH 8.5), 140 µl of 10% ammonium persulfate, 6 µl TEMED and 4.76 ml H_2O. Samples for native PAGE were diluted with 5 x protein loading buffer containing 62.5 µl of 1 M Tris-HCl (pH 8.5), 200 µl glycerol, 687 µl H_2O and 10 µl of 1% bromophenol blue. Native PAGE gels were running in 1 x running buffer, which was diluted from 5 x Native PAGE running buffer (15.1 g Tris-base, 94 g glycine and 1 L H_2O). 5 µl protein mixed with 5 µl of 5 x loading buffer was loaded into 12% Native PAGE, which was running in cold room at 200 V for 60–70 min. 7.5 µg bovine serum albumin protein in GF buffer was used as a marker. The staining and destaining procedures are the same to SDS-PAGE protocol.

2.8.4 Analytical gel-filtration chromatography

For analysis of the multimeric state of Rru-EncFtnSH proteins by analytical gel filtration 25 µl of 90 µM protein was loaded into Superdex 200 PC 3.2/30 column (GE Healthcare Life Sciences) at 15°C with GF buffer running at 0.05 ml/min and pressure limit 0.45 MPa. In order to use analytical gel-filtration (AGF) to determine how metal ions influence the assembly of EncFtn, 90 µM Rru-EncFtnSH monomer fractions were mixed with equal molar concentrations of metal ion solutions including $FeSO_4$ in 0.1% (v/v) HCl, $Fe(NH_4)_2(SO_4)_2$, $FeCl_3$, $CoCl_2$, calcium acetate (CaAc), $ZnSO_4$ and $MnCl_2$ at room temperature for 2 hrs before analysis by AGF. 90 µM Rru-EncFtnSH decamer fractions were loaded to the AGF column as a control for protein assembly. Both monomer and decamer fractions of Rru-EncFtnSH incubated at room temperature for 2 hrs, or overnight, were also analysed as controls to show the stability of the protein samples in the absence of additional metal ions.

2.9 Protein concentration determination

2.9.1 Extinction coefficient calculation

Protein concentration were evaluated using A_{280} extinction coefficients calculated by ExPASy ProtParam (Gasteiger et al. 2005). Briefly, ProtParam tool website (http://web.expasy.org/protparam/) uses protein sequence uploaded to calculate the extinction coefficient in units of $M^{-1}{\cdot}cm^{-1}$ and Abs 0.1% (=1 mg/ml). Protein absorbance at 280nm (A_{280}) was measured using a Thermo Fisher Scientific NanoDrop 2000, or JASCO V-550 UV-Vis Spectrophotometer. Protein concentration in the unit of mg/ml were calculated using the A_{280}/Abs 0.1% extinction coefficient. Protein

concentration in the unit of mol/l were calculated using the A_{280} / ($M^{-1} \cdot cm^{-1}$) extinction coefficient (Simonian & Smith 2001).

2.9.2 BCA microplate assay

Protein concentrations determined by the BCA microplate assay were carried out according to the manufacturer instructions of the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) using the microplate procedure. Briefly, 25 µl of protein sample or Bovine Serum Albumin (BSA) standard (25–1000 µg/ml) was pipetted into a 96-well plate (Greiner) in three technical repeats. BCA working solutions were prepared by mixing A and B solutions in 50:1 ratio; 200 µl of the working solution was added to each well with proteins or BSA standard and mixed on a shaker for 30 s before covering plate and incubating at 37°C for 30 min. A control experiment was performed with 200 µl working solution with 25 µl GF buffer in three repeats. The plate was cooled to room temperature prior to absorbance measurement at 562 nm (A_{562}) on a microplate reader SpectraMax M5 (Molecular Devices). BSA standards were prepared in **Table 2.9.1**. The standard curve was plotted using normalised A_{562} against protein concentration (µg/ml).

BSA concentration (µg/ml)	BSA diluted in GFB	
	GFB buffer (µl)	BSA (ul)
1500	75	225 µl * 2 mg/ml
1000	150	150 µl * 2 mg/ml
750	75	75 µl * 1.5 mg/ml
500	150	150 µl * 1 mg/ml
250	75	75 µl * 500 ug/ml
150	135	15 µl * 1.5 mg/ml
100	135	15 µl * 1 mg/ml
50	135	15 µl * 500 ug/ml
25	135	15 µl * 250 ug/ml
0	135	0

Table 2.9.1 Bovine Serum Albumin standards.

Bovine Serum Albumin (BSA) protein standards 1.5 mg/ml and 1 mg/ml were prepared by diluting the 2 mg/ml commercial stock. The 750, 500, 150 and 100 µg/ml BSA samples were prepared by diluting the 1.5 or 1 mg/ml samples. 250 and 50 µg/ml samples were made from the dilution of 500 µg/ml sample. The 25 µg/ml sample was prepared from the dilution of 250 µg/ml.

2.10 Multiple sequence alignment and phylogenetic tree construction

The ferritin superfamily contains 24-meric mammalian ferritins (H-/M-/L-chains), bacterial ferritin (FTN), bacterioferritin (BFR) and erythrin domain of rubrerythrin protein (ER), 12-meric mini-ferritin (DPS) and ferritin-like proteins with shorter sequences. The multiple protein sequence alignment of ferritin superfamily was initially carried out using programs such as Clustal Omega (Sievers et al. 2011; Li et al. 2015; McWilliam et al. 2013) and PROMALS3D (Pei et al. 2008). The alignment of the conserved ferroxidase centres among ferritin family proteins except DPS and L-chain ferritin has been used to estimate the alignment quality. However, because of the diverged ferritin classes, the one-step alignment failed to produce an aligned ferroxidase centre. Therefore, a progressive alignment was performed by setting up local alignment first. Mammalian ferritins were aligned using T-coffee (Notredame et al. 2000) or Clustal Omega as the seed of alignment. Then FTN family was aligned using T-coffee, which was further added to the aligned mammalian ferritins and aligned by MAFFT (Katoh et al. 2002). For evolutionarily diverged proteins the matrix BLOSUM 45 was used in this study. The other ferritin families including BFR, ER, DPS and ferritin-like proteins were added in a step-wise manner and aligned by MAFFT gradually. Apart from the conserved ferroxidase centre, the C-terminal localisation sequences of encapsulin-associated ferritin-like proteins (EncFtn) were also conserved, which has been taken into account when assessing the alignment quality. To rule out the bias introduced by the alignment order, the order of progressive alignment has been altered, which did not affect the alignment of both FOC and C-terminal location sequence of EncFtn. The phylogenetic tree was built in MEGA7 using Neighbour-joining method. 500 of bootstrap replications were applied in the phylogeny test. p-distance model (Nei & Kumar 2000)was used in the substitution model for amino acid substitutions. The rate of variation among sites was modelled with a gamma distribution (shape parameter = 2.5). The analysis involved 104 amino acid sequences. All ambiguous positions were removed for each sequence pair (Kumar et al. 2016).

2.11 The modelling of encapsulin nanocompartments

The modelling of encapsulin nanocompartments was carried out in PyMOL manually. Firstly, *T. maritima* encapsulin (Tm-Enc) was loaded into PyMOL 1.8 as a 60-mer using 'PDB analysis-Assemblies' in the plugin. Secondly, Rru-EncFtn was loaded into PyMOL and a decamer was docked on top of a pentamer of Tm-Enc along its 5-fold symmetry axis using 'translate' and 'rotate' related command lines. Once a decamer of Rru-EncFtn was docked to Tm-Enc, the C-terminal position of Rru-EncFtn should be adjacent to the localisation peptide found in Tm-Enc structure (PDB ID: 3DKT). This docking was repeated 12 times to constitute a full encapsulin nanocompartment. This manual docking process was relatively labour-intensive and could possibly give rise to clashes in the protein-protein interface. Another docking program 'HADDOCK' can be introduced as a tool to validate this manual docking result (Van Zundert et al. 2016). In addition, the protein-protein interaction between a decamer of Rru-EncFtn and another decamer, or between a decamer of Rru-EncFtn and its docking pentamer of Tm-Enc, can be evaluated using PDBePISA server (Krissinel & Henrick 2007); this can be an independent estimation of manual docking and HADDOCK.

Chapter 3 Purification, structure, and assembly of *R. rubrum* EncFtn

3.1 Introduction

R. rubrum EncFtn (KEGG ID: Rru_A0973) shares 57% protein sequence identity with protein NE0167 (PDB ID: 3K6C) from *N. europaea* (Chang et al. n.d.). The C-terminus of NE0167 is located outside its ring-like decameric structure, suggesting that C-terminal his-tag of *R. rubrum* EncFtn (Rru-EncFtn) may not interfere with the structure assembly. Based on the sequence alignment of Rru-EncFtn and NE0167, a construct with the first 96 amino acids of Rru-EncFtn and a C-terminal hexahistidine tag (Rru-EncFtnSH) was designed (**Figure 3.1.1**). A full-length gene *Rru_A0973* was inserted into the pACYCDuet-1 vector and expressed in *E. coli* as Rru-EncFtn; and the truncated gene *Rru_A0973*$_{1\text{-}96}$ was inserted into pET-28a(+) and expressed with C-terminal his-tag as Rru-EncFtnSH. Both constructs were overexpressed in *E. coli* cells. Rru-EncFtnSH was further characterised in two aspects: protein assembly and crystal structure.

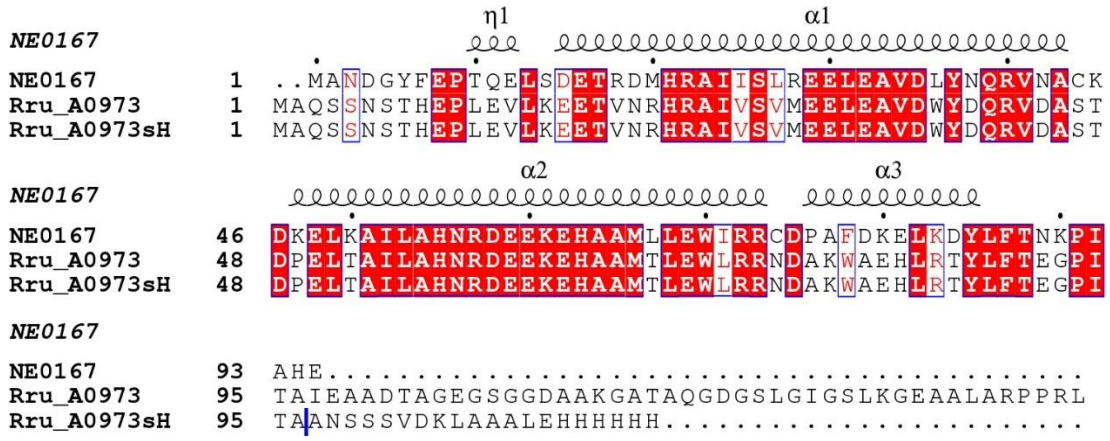


Figure 3.1.1 Protein sequence alignment of Rru-EncFtn and NE0167.

A sequence alignment of Rru-EncFtn (Rru_A0973) and NE0167 was performed using Clustal Omega (McWilliam et al. 2013; Sievers et al. 2011) and edited with Esprit 3.0 (Robert & Gouet 2014). The secondary structure of protein NE0167 (PDB ID: 3K6C) is shown above each row of the alignment. The Rru-EncFtnSH protein (Rru_A0973sH) contains 1–96 amino acids of Rru-EncFtn and a C-terminal hexa-histidine tag (after the blue vertical line).

3.2 Protein purification

3.2.1 Calibration of HiLoad 16/600 Superdex 200pg gel-filtration column

Size-exclusion gel-filtration chromatography provides a means to estimate the molecular weight or size (Stokes radius) of proteins under certain buffer and temperature conditions. The molecular weight of unknown proteins can be determined by comparing the elution volume related parameter K_{av} to those calculate from known calibration standards run on the same column. Here a calibration curve for a HiLoad 16/600 Superdex 200pg column was prepared by running LMW/HMW calibration kits (GE Healthcare) to measure the elution volume (V_e) of each protein standard (**Figure 3.2.1 A**). The void volume (V_o) of S200 column was estimated based on the retention volume of blue dextran 2000. The K_{av} value was derived based on the equation:

$$K_{av} = (V_e - V_o) / (V_t - V_o)$$

Where V_t equals to the total bed volume of particular column. K_{av} was then plotted against the logarithm of molecular weight of protein standards (**Figure 3.2.1 B**). It should be borne in mind that protein shape also affects the K_{av}-based molecular weight estimation and the assumption that the protein samples are approximately spherical.

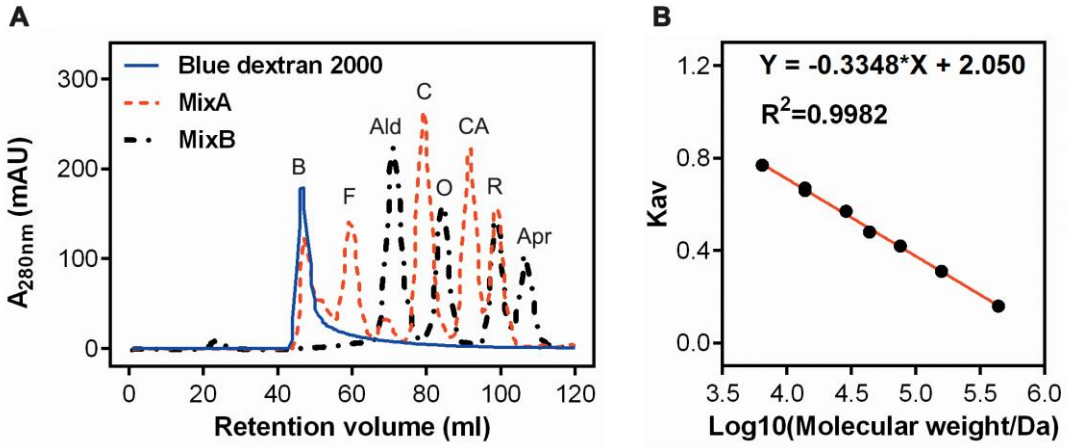


Figure 3.2.1 Chromatographic separation and calibration curve for HiLoad 16/600 Superdex 200 pg column.

(A) Chromatogram of HiLoad 16/600 Superdex 200 pg gel-filtration column calibration using blue dextran 2000 (B, blue solid curve), MixA (red dashed curve) and MixB (black dashed curve) in which blue dextran determines the void volume. MixA contains ferritin (F, 440 kDa), conalbumin (C, 75 kDa), carbonic anhydrase (CA, 29 kDa) and ribonuclease A (R, 13.7 kDa). MixB consists of ovalbumin (O, 44 kDa), aprotinin (Apr, 6.5 kDa), ribonuclease A and aldolase (Ald, 158 kDa). (B) Calibration curve of HiLoad 16/600 Superdex 200 pg column.

3.2.2 Rru-EncFtn purification

The recombinant Rru-EncFtn protein possesses 140 amino acids with a predicted molecular weight of 15,192.74 Da, and an isoelectric point (pI) of 4.75 (Gasteiger et al. 2005); this untagged protein was purified using anion exchange chromatography (AEC) with a 12 ml Q Sepharose™ Fast Flow ion exchange column (**Figure 3.2.2**).

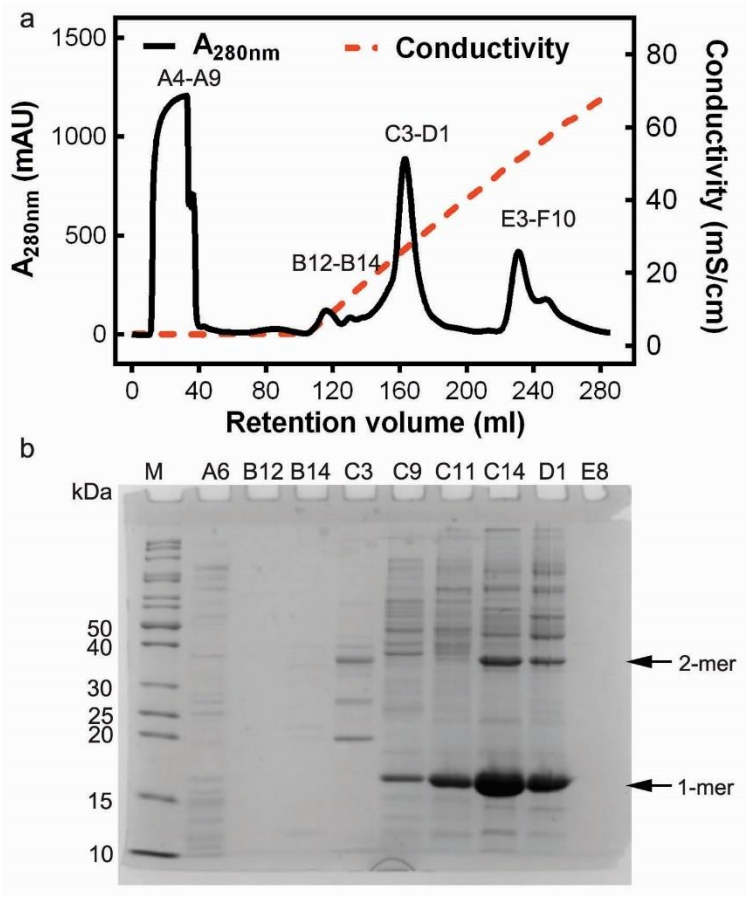


Figure 3.2.2 Anion exchange chromatography purification of Rru-EncFtn using a Q sepharose column.

(A) Q Sepharose purification chromatogram of Rru-EncFtn. Fractions from A4 to A9 were flow-through. Proteins were eluted by a linear gradient over 20 column volumes from 0–1 M NaCl as fractions from B12 to F10. (B) Fractions from (A) were analysed by 15% SDS-PAGE, in which only fractions C9–D1 showed a band at 15 kDa, corresponding to the monomer size of Rru-EncFtn. The monomer and dimer sized bands were highlighted by black arrows.

Fractions with the protein of interest, assessed by the presence of bands at an estimated molecular weight of 15 kDa on 15% SDS-PAGE were pooled and concentrated to 2 ml prior to Size-Exclusion Chromatography (SEC) using a HiLoad 16/600 Superdex 200 pg column (**Figure 3.2.3**). The SEC purification chromatogram of Rru-EncFtn displayed two major peaks at around 62 ml and 82 ml, respectively (**Figure 3.2.3**).

Fractions analysed by 15% SDS-PAGE showed that the first peak (fractions C5–C7) showed two major bands at around 15 kDa and 30 kDa, corresponding to the approximate monomer and dimer size of full-length Rru-EncFtn. The second peak (fractions C12–C15) showed a major band at around 15 kDa, consistent with the monomer size. These bands were excised from the gel and subjected to MALDI-TOF peptide mass fingerprinting (PMF), and were all identified as the *R. rubrum* ferritin-like protein, Rru-EncFtn. It is noteworthy that there is no cysteine in Rru-EncFtn protein, it is unclear what form of the protein correlates with the dimer size band on SDS-PAGE. According to the calibration curve of the SEC column (**Figure 3.2.1**) and SDS-PAGE analysis, the two chromatographic peaks indicate the presence of different oligomeric states of Rru-EncFtn during the SEC purification.

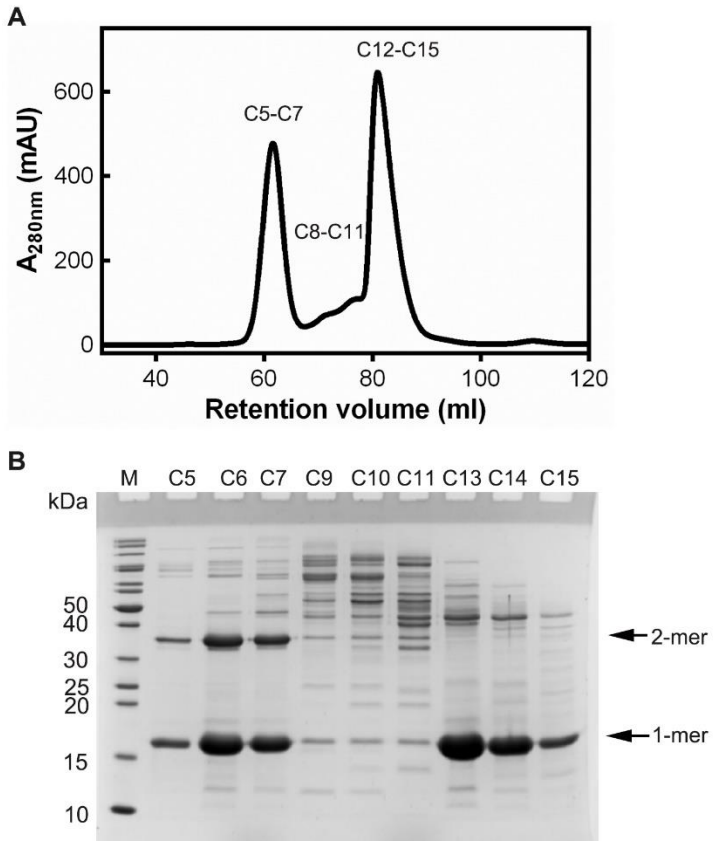


Figure 3.2.3 Size exclusion chromatography purification of Rru-EncFtn using HiLoad 16/600 Superdex 200 pg column.

(A) Gel-filtration chromatogram of Rru-EncFtn. Proteins were eluted by GF buffer into fractions C5–C15, of which two major chromatogram peaks were found at 62 ml and 82 ml. (B) 15% SDS-PAGE analysis of fractions C5–C15. The C5–C7 peak in (A) corresponds to two major bands around 15 and 30 kDa while the C12–C15 peak in (A) is congruous with a main band around 15 kDa.

The SEC purification of Rru-EncFtn were repeated several times from different batches of *E. coli* grown in LB medium (**Figure 3.2.4**). Each purification run displayed two major chromatographic peaks around 62 and 82 ml but with distinct peak ratios. It indicated that alteration of oligomeric states of Rru-EncFtn might be a result of different levels of trace elements in the different batches of LB media used in these experiments.

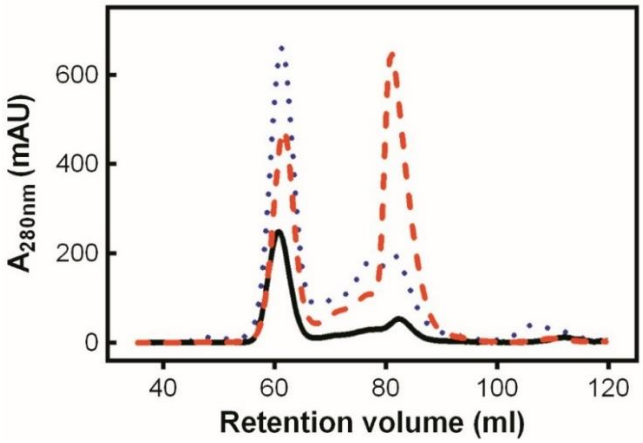


Figure 3.2.4 Overlay of size exclusion chromatograms of Rru-EncFtn purified by SEC HiLoad 16/600 Superdex 200 pg column.

Rru-EncFtn purification from *E. coli* BL21(DE3) grown in three batches of LB media showed different gel-filtration profiles in black solid, red dashed and blue dotted curves. They all displayed two major peaks around 62 and 82 ml but with different ratios of peak areas.

3.2.3 Rru-EncFtnSH purification

Rru-EncFtnSH (Rru-EncFtn$_{1–96}$ with C-terminal hexa-histidine tag) contains 116 amino acids and has a molecular weight of 13,325.67 Da and a pI of 5.34 as predicted by ExPASy ProtParam tool (http://web.expasy.org/protparam/) (Gasteiger et al. 2005). The protein was purified by IMAC and showed two major bands around 13 kDa and 27 kDa from eluted fractions, corresponding to the monomer and dimer size of Rru-EncFtnSH, respectively (**Figure 3.2.5**).

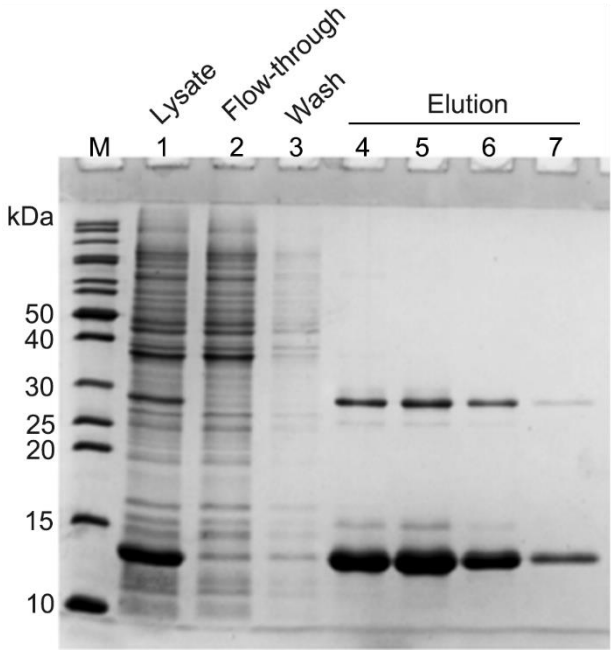


Figure 3.2.5 15% SDS-PAGE analysis of IMAC purified Rru-EncFtnSH.

Rru-EncFtnSH protein was purified using a 5 ml HisTrap FF column manually and subjected to SDS-PAGE analysis. Lane M: protein ladder; Lane 1: cell lysate; lane 2: flow-through; lane 3: unbound proteins washed by hisA buffer; lane 4–5: protein eluted by 50% hisA mixed with 50% hisB buffer; lane 6–7: protein eluted by 100% hisB buffer.

Fractions with protein of interest were pooled and concentrated to 2 ml for SEC purification (**Figure 3.2.6**). The Rru-EncFtnSH displayed higher purity than Rru-EncFtn after two-step purification due to the presence of His-tag. The S200 SEC purification of Rru-EncFtnSH displayed three chromatographic peaks. Peak1 (fractions A2–A7) was consistent with the void volume of the column (46 ml). Since the S200 column has the separation range between 10–600 kDa, peak1 is likely to be aggregated material, or high-order oligomers of the protein. Peak2 (Fractions A11–B9) separated as a mixture of the 13 kDa and 26 kDa bands by SDS-PAGE, consistent with the monomer and dimer mass of Rru-EncFtnSH, respectively. Peak3 (Fractions B11–

C5) resolved as a 13 kDa band by SDS-PAGE, with a minor band around 26 kDa. These bands at 13 kDa and 26 kDa were both identified as Rru-EncFtn by PMF. According to the disassembly experiment in **Figure 3.5.5**, the dimer bands seen on the SDS-PAGE might be the dimer with metal association, which will be discussed later in **Section 3.5.3**.

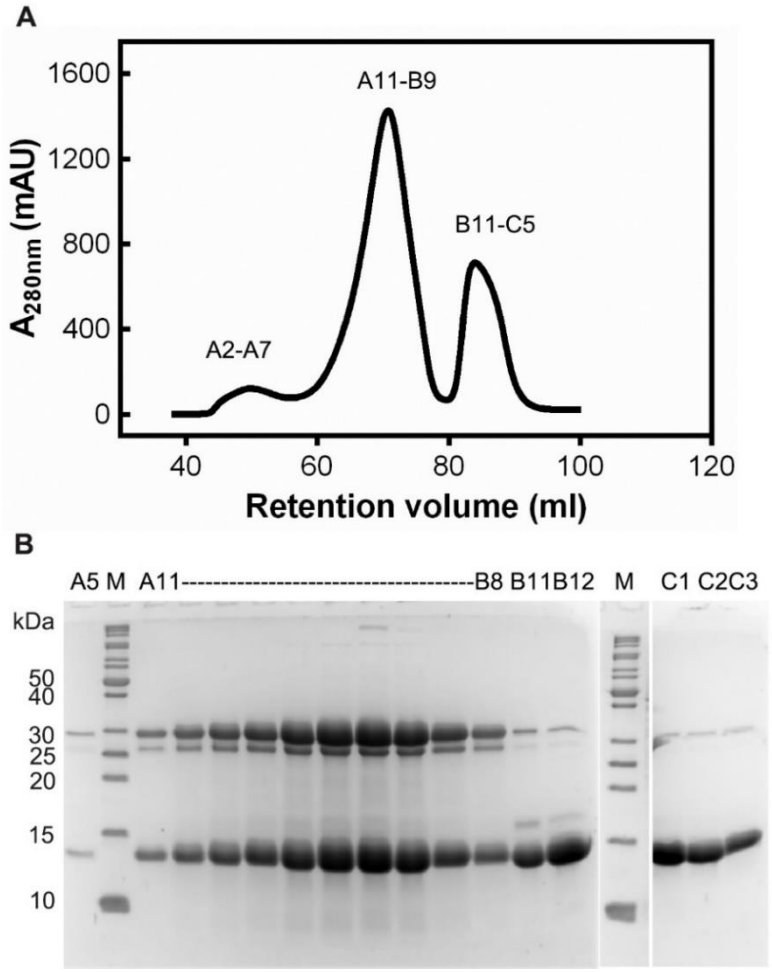


Figure 3.2.6 Size exclusion chromatography purification of Rru-EncFtnSH using HiLoad 16/600 Superdex 200 pg column.

(A) Gel-filtration chromatogram of Rru-EncFtnSH. Proteins were eluted using GF buffer into fractions A2–C5, of which three chromatogram peaks are shown. (B) 15% SDS-PAGE analysis of fractions A2–C5. Three peaks of (A) correspond to a mixture of dimer size band and a monomer size band at different ratios.

3.3 Oligomeric states of Rru-EncFtnSH

The Rru-EncFtnSH SEC purification from LB medium displayed a similar chromatographic pattern with Rru-EncFtn. To improve between-batch consistency and further explore the role that trace elements play in the oligomeric state of this protein, further protein production experiments of Rru-EncFtnSH were performed using the defined M9 minimal medium.

3.3.1 Rru-EncFtnSH purified from MM ± Fe(II)

To investigate if it is traces of iron ions in LB medium influence protein oligomeric states, Rru-EncFtnSH was purified from *E. coli* BL21(DE)3 grown in minimal media (MM) and MM supplemented with 1 mM $Fe(NH_4)_2(SO_4)_2$ (dissolved in 0.1% (v/v) HCl) (MM + Fe(II)), respectively. **Figure 3.3.1** showed the superimposed SEC chromatograms of Rru-EncFtnSH. The Rru-EncFtnSH purified from MM displayed a major peak FP2 and a negligible FP1 peak, with higher-order oligomers; while, more high-order oligomers of Rru-EncFtnSH formed when produced in *E. coli* grown in MM + Fe(II), indicating that the iron content in the medium influences the assembly of Rru-EncFtnSH.

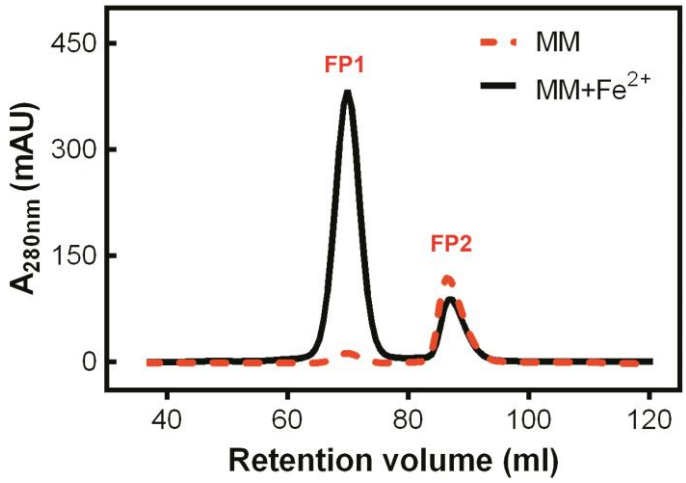


Figure 3.3.1 Overlay of SEC chromatograms of Rru-EncFtnSH purified from MM with and without $Fe(NH_4)_2(SO_4)_2$.

Rru-EncFtnSH purified from MM ± Fe(II) displayed two major chromatographic peaks in different ratios, consistent with the purification profile from LB medium. The higher proportion of oligomer in the left peak is seen in the sample purified from MM + Fe(II), suggesting that Fe(II) facilitates the oligomerisation of Rru-EncFtnSH. Fractions from two chromatographic peaks of MM were pooled as FP1 and FP2, respectively.

3.3.2 Determination of Rru-EncFtnSH oligomeric states by SEC-MALS

To more rigorously estimate the molecular weight of the oligomeric complexes of the recombinant Rru-EncFtnSH, samples purified from MM were analysed by size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). Experiments were performed using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in GF buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl). Rru-EncFtnSH peak fractions (FP1 and FP2) purified from protein produced in MM (shown in **Figure 3.3.1**) were pooled separately and concentrated to 2–4 mg/ml prior to SEC-MALS analysis (**Figure 3.3.2**). Two monodispersed peaks were discovered in both FP1 and FP2, suggesting an equilibrium between the two oligomeric states. FP1 was determined as a major decamer (132 kDa) with a small proportion of monomer (15 kDa). FP2 consisted of a major monomer (13 kDa) and minor proportions of decamer (126 kDa). These SEC-MALS results are consistent with the observation that FP1 consisted of mainly decamer of Rru-EncFtnSH while FP2 was mainly the monomer of Rru-EncFtnSH.

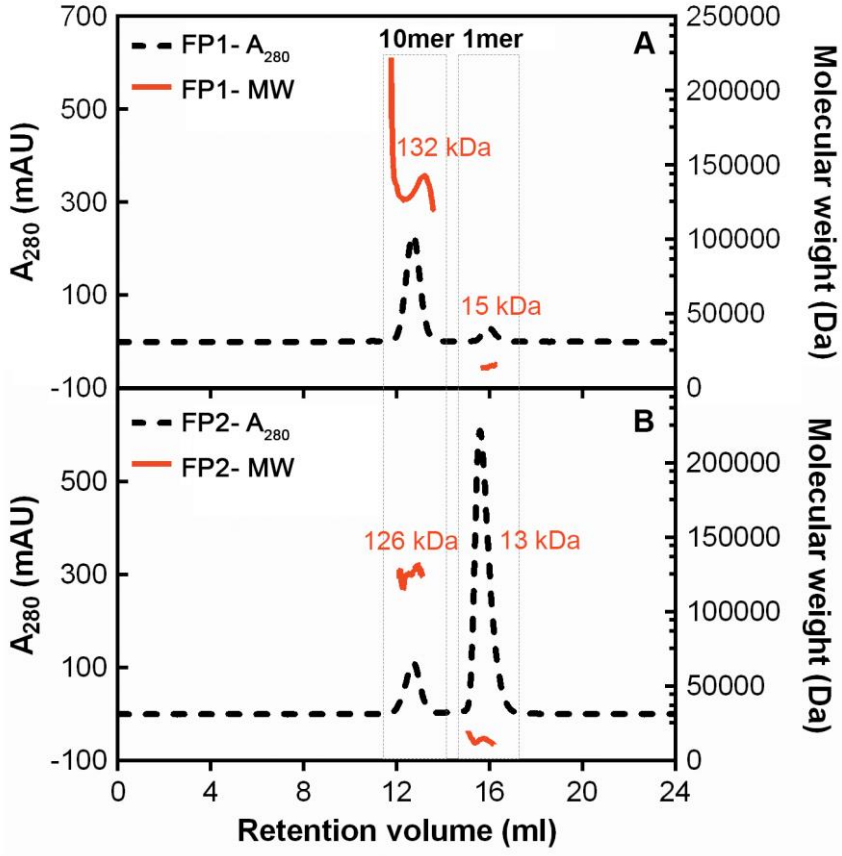


Figure 3.3.2 SEC-MALS analysis of Rru-EncFtnSH oligomeric states on a Superdex 200 10/300 GL column.

Red solid curve indicates the calculated molecular weight (MW) of Rru-EncFtnSH fractions. Black dashed curve indicates A_{280}. (A) Sample Rru-EncFtnSH-FP1 showed two peaks, corresponding to 132 kDa and 15 kDa, respectively. (B) Sample Rru-EncFtnSH-FP2 showed a decamer peak (126 kDa) and a monomer peak (13 kDa). The molecular weight of the major peak (decamer) in FP1 sample was consistent with the theoretic decamer size (133 kDa); the MW of major peak (monomer) in FP2 was equivalent to the theoretic monomer size (13 kDa), indicating that Rru-EncFtnSH appeared as a mixture of decamer and monomer in solution.

3.3.3 Metal quantification by inductively coupled plasma mass spectrometry

Given the observation that the presence of iron ions in medium affected the protein assembly of Rru-EncFtnSH, the iron binding of purified Rru-EncFtnSH was analysed by Inductively Coupled Plasma with Mass Spectrometry (ICP-MS). Fractions of SeMet-labelled Rru-EncFtnSH purified from SEC were subjected to ICP-MS analysis by Dr. Emma Tarrant and Dr. Kevin Waldron (Institute for Cell and Molecular Biosciences, Newcastle University). ICP-MS is a technique that can quantify most

elements to different detection limits. Although the measurement of both S and Se by ICP-MS are problematic due to the interference of dioxygen ($^{16}O_2$) on ^{32}S and $^{40}Ar_2$ on ^{80}Se; the detection of Se proves to be easier in practice. That is why the SeMet-labelled Rru-EncFtnSH were used to estimate the protein concentration in ICP-MS instead of the native protein.

The plasmid pET-28a-Rru-EncFtnSH was transformed into methionine auxotrophic *E. coli* B834(DE3) cells, which were incubated in SeMet medium to allow the substitution of methionine by the analogue SeMet during protein synthesis. 1 mM $Fe(NH_4)_2(SO_4)_2$ was supplemented to the SeMet medium to ensure the iron incorporation of Rru-EncFtnSH. The SeMet-labelled Rru-EncFtnSH was then purified by IMAC and SEC. The SEC chromatogram along with the SDS-PAGE analysis showed results consistent with previous purifications, with a high proportion of decamer peak and a low abundance of monomer peak in the presence of iron in the SeMet medium (**Figure 3.3.3**), suggesting that the selenium labelling did not alter the assembly of Rru-EncFtnSH. The A_{315} was also monitored to indicate the higher iron content in decamer fractions whereas in monomer fractions the A_{315} was undetected (Bonomi et al. 1996) (**Figure 3.3.3**). Each fraction was subjected to ICP-MS analysis, of which 100 times more iron was detected in the decamer peak fraction than the monomer fraction (**Figure 3.3.3 A** & **Table 3.3.1**).

The SeMet incorporation rate was determined as 100% through LC-MS by Dr. David Clarke (School of Chemistry, University of Edinburgh). The LC-MS result also confirmed that the recombinant Rru-EncFtnSH sequence only contains two methionines due to the N-terminal methionine being excised by the methionine aminopeptidase *in vivo* (Frottin et al. 2006). Therefore, the Rru-EncFtnSH concentration can be derived as half of the Se concentration detected by ICP-MS. The sub-stoichiometric iron binding of Rru-EncFtnSH was deduced based on ICP-MS analysis (**Table 3.3.1**), which is in contrast to the classical ferritins (Le Brun et al. 2010).

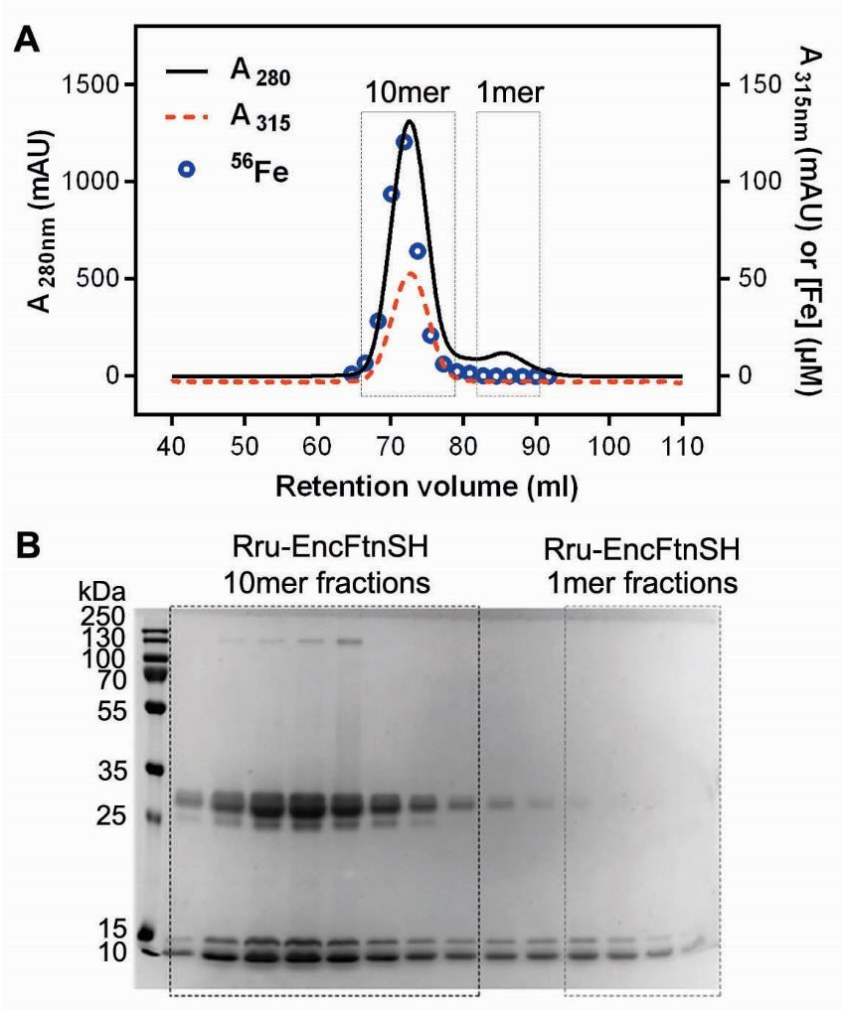


Figure 3.3.3 SEC purification of SeMet-labelled Rru-EncFtnSH with Fe content detected by A_{315} and ICP-MS.

(A) SEC purification chromatogram of SeMet-labelled Rru-EncFtnSH using HiLoad 16/600 Superdex 200 pg column. Protein concentration was monitored by A_{280} (black curve). Fe content was detected by A_{315} (red dashed curve). Blue circles stands for the Fe content of each fraction (B3–C5) analysed via ICP-MS by Dr. Emma Tarrant and Dr. Kevin Waldron (ICMB, Newcastle University). (B) 18% SDS-PAGE analysis of fractions in (A). The decamer fractions displayed two major bands around 13 kDa and 26 kDa as well as a minor band of 130 kDa in the peak fractions. The monomer fractions present mainly 13 kDa bands, correponding to monomeric Rru-EncFtnSH. ICP-MS data coupled with A_{315} show that Fe content differs in decamer and monomer of Rru-EncFtnSH.

Peak	Rru-EncFtnSH elution volume (ml)	Element concentration (μM)				Derived Rru-EncFtnSH concentration (μM)	Derived Fe/Rru-EncFtnSH monomer
		Ca	Fe	Zn	Se		
10mer	66.49	n.d.	6.67	n.d.	24.60	12.30	0.54
	68.29	n.d.	28.41	0.003	124.49	62.25	0.46
	70.09	2.85	93.68	2.43	301.73	150.87	0.62
	71.89	6.85	120.59	3.65	379.76	189.88	0.64
	73.69	1.85	64.41	0.79	240.58	120.29	0.54
	75.49	0.9	21.12	n.d.	101.65	50.82	0.42
	77.29	n.d.	6.21	n.d.	42.63	21.31	0.29
	79.09	0.1	2.43	n.d.	26.54	13.27	0.18
	80.89	0.95	1.50	n.d.	22.30	11.15	0.14
	82.69	n.d.	0.20	n.d.	29.20	14.60	0.01
1mer	84.49	n.d.	0.10	n.d.	34.94	17.47	0.006
	86.29	n.d.	0.004	n.d.	28.85	14.43	n.d.
	88.09	n.d.	n.d.	n.d.	17.35	8.67	n.d.
	89.89	n.d.	n.d.	n.d.	5.54	2.77	n.d.
	91.69	n.d.	n.d.	n.d.	0.13	0.07	0.21

Table 3.3.1 ICP-MS data of SeMet-labelled Rru-EncFtnSH fractions purified from SEC.

Rru-EncFtnSH was purified as a SeMet derivative from *E. coli* B834(DE3) cells grown in SeMet medium supplemented with 1 mM $Fe(NH_4)_2(SO_4)_2$. Fractions from size-exclusion gel filtration were collected, acidified and analysed by ICP-MS. Rru-EncFtnSH concentration was calculated based on the presence of two SeMet per mature monomer. Samples where the element was undetectable are labelled with n.d..

Taken together, both A_{315} and ICP-MS results suggested that the decamer and monomer of Rru-EncFtnSH differ in iron content. These results merited further investigation on questions as follows: (1) can the oligomeric state of Rru-EncFtnSH be altered in the presence of different metals *in vitro*? (2) Does Rru-EncFtnSH have a preference for a particular oxidation state of iron? To answer these questions, Rru-EncFtnSH monomer fractions were pooled as iron-free samples for later experiments.

3.4 Structure analysis of Rru-EncFtnSH

3.4.1 Crystallisation and structure determination of Rru-EncFtnSH

Rru-EncFtnSH was purified from *E. coli* BL21(DE3) cells grown in LB medium using Q Sepharose AEC to minimise the interference of nickel carried over from an IMAC column, followed by SEC purification. The monomer and decamer fractions were pooled separately and concentrated to around 11 mg/ml before crystallisation trials using commercial kits described in **Section 2.5.1**. Rru-EncFtnSH crystals were initially found in the PEG/Ion-C4 screen solution (Hampton Research), which includes 0.2 M calcium acetate hydrate (CaAc) and 20% (w/v) PEG 3350. The crystallisation of Rru-EncFtnSH was further optimised in a mixture of 14% (w/v) PEG 3350 and 0.15 M calcium acetate (CaAc) using hanging-drop vapour diffusion. Rod-like crystals of 0.2–0.3 mm in length formed in around five days; these were then transferred into a cryo-protectant solution containing 14% (w/v) PEG 3350, 0.15 M CaAc and 20% (v/v) PEG 200, before flash cooling in liquid nitrogen. Both monomer and decamer fractions crystallised and diffracted to slightly different resolution limits as 2.0 Å and 2.3 Å, respectively, of which the X-ray diffraction datasets of monomer fractions were collected at beamline I04 (Diamond Light Source) and used in the structure determination for the best data statistics. The structure of Rru-EncFtnSH was determined by molecular replacement in space group $P\ 1\ 2_1\ 1$, to 2.0 Å resolution, using PDB ID: 3K6C as the search model (see **Table 3.4.1** for X-ray data collection and refinement statistics). All structure figures were generated using PyMOL 1.8 (Schrodinger 2015).

	Rru-EncFtnSH
Data collection	
Beamline	I04
Wavelength (Å)	1.73
Resolution range (Å)	49.63–2.06 (2.10–2.06)
Space group	*P* 1 2_1 1
Unit cell (Å) *a*	98.18
b	120.53
c	140.3
β (°)	95.36
Total reflections	1,264,922 (41,360)
Unique reflections	197,873 (8,766)
Multiplicity	6.4 (4.7)
Anomalous multiplicity	3.2 (2.6)
Completeness (%)	99.2 (88.6)
Anomalous completeness (%)	96.7 (77.2)
Mean I/sigma(I)	10.6 (1.6)
Wilson B-factor	26.98
R_{merge}	0.123 (0.790)
R_{meas}	0.147 (0.973)
CC1/2	0.995 (0.469)
CC*	0.999 (0.846)
Image DOI	10.7488/ds/1342
Refinement	
R_{work}	0.171 (0.318)
R_{free}	0.206 (0.345)
Number of non-hydrogen atoms	23,222
macromolecules	22,276
ligands	138
water	808
Protein residues	2,703
RMS(bonds) (Å)	0.012
RMS(angles) (°)	1.26
Ramachandran favored (%)	100
Ramachandran allowed (%)	0
Ramachandran outliers (%)	0
Clash score	1.42
Average B-factor (Å^2)	33.9
macromolecules	33.8
ligands	40.4
solvent	36.2
PDB ID	5DA5

Table 3.4.1 Rru-EncFtnSH X-ray data collection and refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses. Friedel mates were averaged when calculating reflection numbers and statistics.

The crystallographic asymmetric unit contained thirty monomers of Rru-EncFtnSH with visible electron density for residues 7–96 in each chain. The protein chains were arranged as three identical annular decamers, each with D5 symmetry. The decamer has a diameter of 7 nm and thickness of 4 nm. The arrangement of ten monomers in alternating orientation forms the decamer of Rru-EncFtnSH, which assembles as a pentamer of dimers. Each monomer lies at 45° relative to the vertical central-axis of the ring, with the N-termini of alternating subunits capping the center of the ring at each end, while the C-termini are arranged around the circumference (**Figure 3.4.1 A–B**). The central hole in the ring is 2.5 nm at its widest in the center of the complex, and 1.5 nm at its narrowest point near the outer surface, although it should be noted that a number of residues at the N-terminus are not visible in the crystallographic electron density and these may occupy the central channel. The surface of the decamer has distinct negatively charged patches, both within the central hole and on the outer circumference, which form spokes through the radius of the complex (**Figure 3.4.1 C–E**).

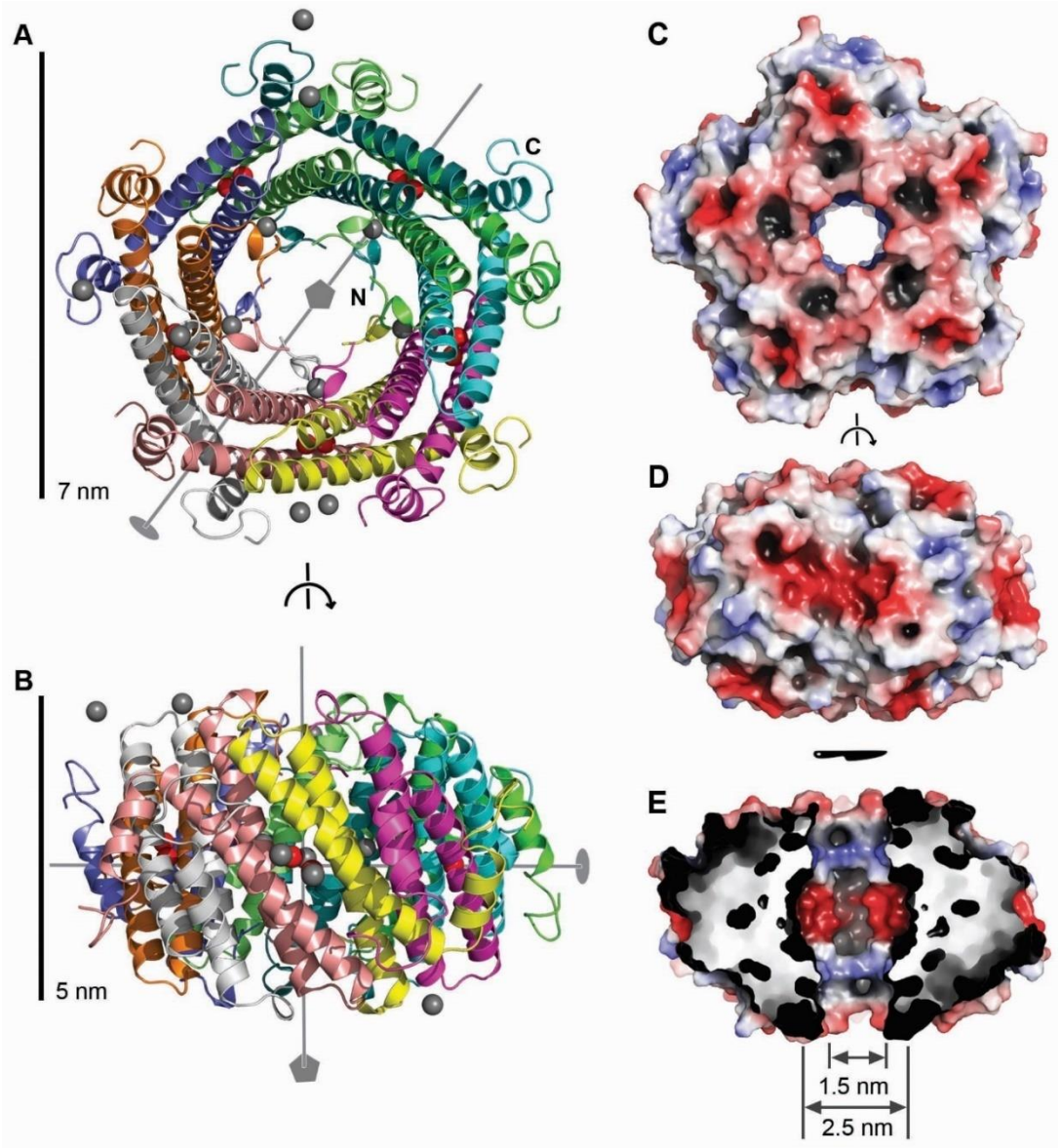


Figure 3.4.1 Ring-like decamer structure of Rru-EncFtnSH.

Top view of (A) Rru-EncFtnSH decamer structure is shown in cartoon mode and coloured by chain with N-termini reaching the central ring and C-termini stretching towards outer surface. Iron ions are shown as red spheres in the dimer interface. Calcium ions are shown in grey spheres bound in dimer interfaces or as surface residues. (B) Side view of decameric Rru-EncFtnSH. (C–E) Electrostatic surface of Rru-EncFtnSH decamer. The solvent accessible surface of Rru-EncFtnSH is shown and coloured by electrostatic potential as calculated using the APBS plugin in PyMOL (Baker et al. 2001). Negatively charged regions are coloured red and positive regions in blue, neutral regions in grey. (C) View of the surface of the Rru-EncFtnSH decamer looking down the central axis. (D) Orthogonal view of (C). (E) Cutaway view of (D) shows the charge distribution within the central cavity. The central channel is 1.5–2.5 nm wide.

3.4.2 Rru-EncFtnSH structure analysis

The Rru-EncFtnSH monomer consists of an N-terminal 3_{10}-helix followed by two 4 nm long antiparallel alpha-helices—α1 (amino acid residues 17–45), α2 (49–76) and a shorter 1.4 nm long C-terminal α3 (78–87) (**Figure 3.4.2 A**); α1 and α2 are arranged with their long axes at 158.63° to each other (**Figure 3.4.2 B**); α3 projects at 114.61° to α2.

Two monomers, arranged in an antiparallel manner, assemble into a FOC interface with metal (Fe and Ca) ions bound by negatively charged residues (**Figure 3.4.3 C**); and a non-FOC interface with neutral-charged residues (**Figure 3.4.3 D**). According to PDBePISA analysis (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) on interface interaction (Krissinel & Henrick 2007), there are no sulphide bridges, or covalent bonds in the structure. The non-FOC interface is stabilised by 20 hydrogen bonds and 14 salt bridges, of which 33% surface area is buried. The FOC interface, burying 17% surface area, consists of 10 hydrogen bonds, 8 salt bridges and a number of metal ions (**Table 3.4.2**). One point to be noted is that the PDBePISA server does not calculate the contribution of metal binding to FOC interface, indicating that the interaction of FOC interface might have been underestimated. The two dimer interfaces will be compared later by the disassembly experiment via collision-induced dissociation (CID) tandem mass spectrometry (**Figure 3.5.5**).

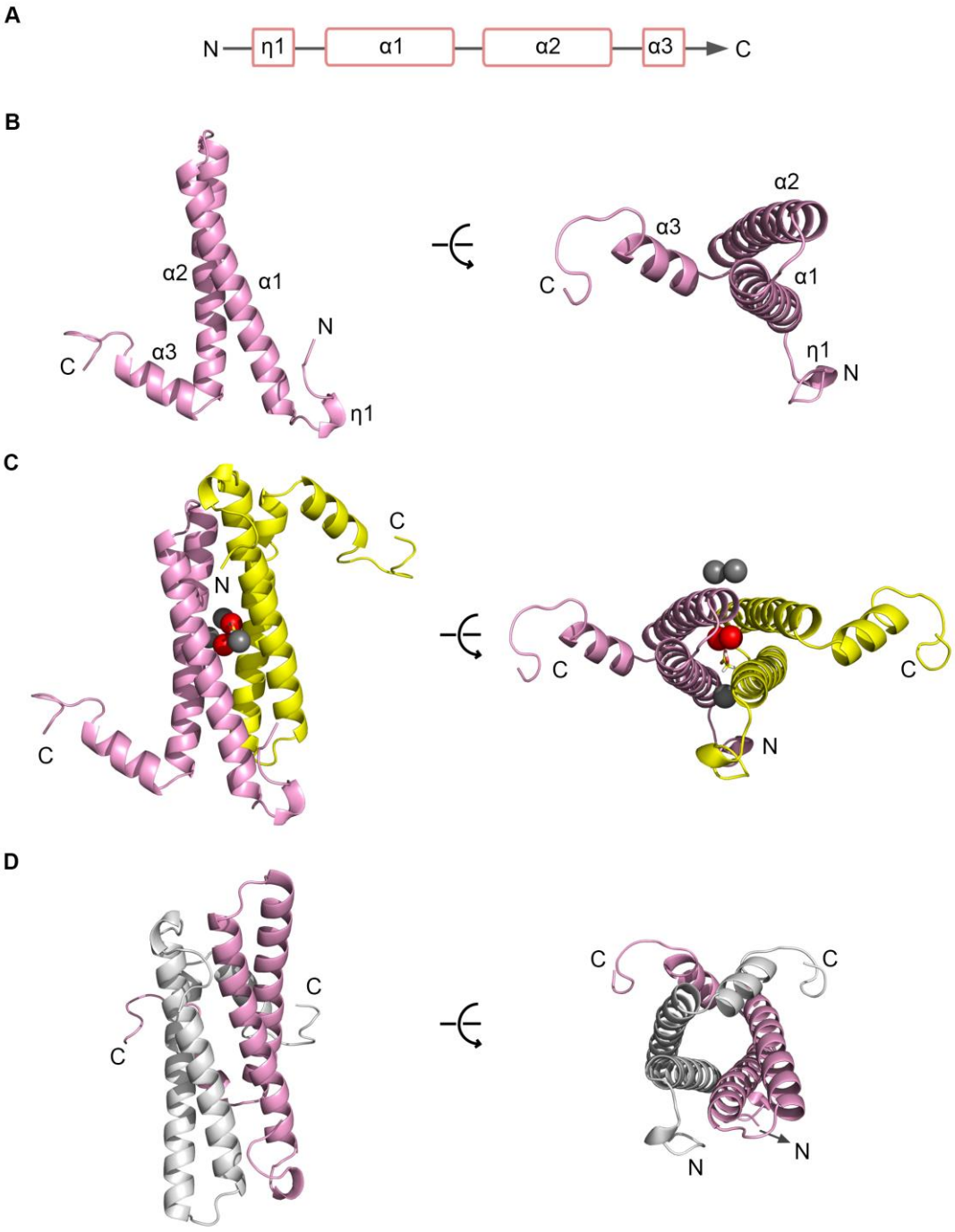


Figure 3.4.2 Rru-EncFtnSH structure analysis.

Rru-EncFtnSH structure is presented in cartoon mode and coloured by chain. (A) Schematic representation of the secondary structure of Rru-EncFtnSH monomer. (B) Monomer structure of Rru-EncFtnSH. (C) Structure of Rru-EncFtnSH dimer consists of four alpha-helical bundles with metal ions bound as red and grey spheres in the interface. (D) Structure of Rru-EncFtnSH dimer without metal binding in the second interface.

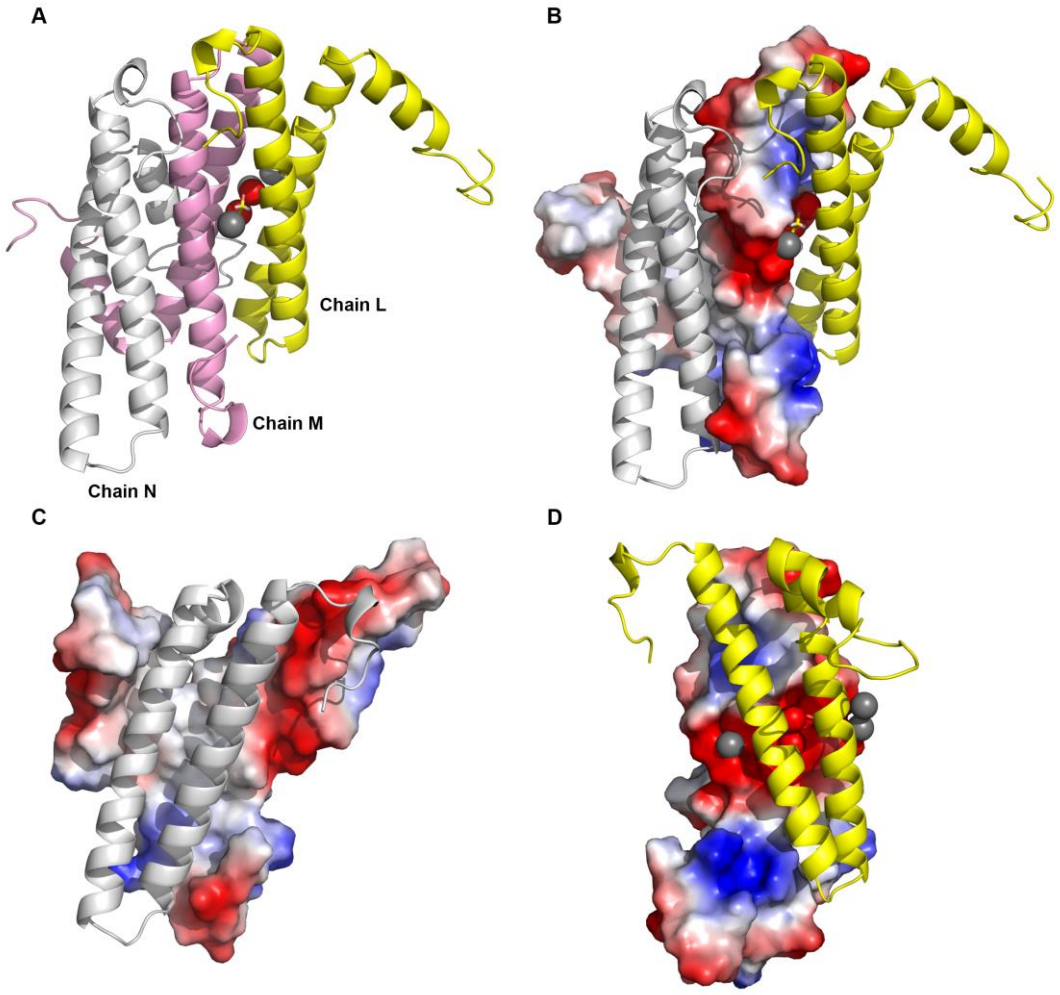


Figure 3.4.3 Rru-EncFtnSH dimer interfaces.

(A) Three chains of Rru-EncFtnSH are shown in cartoon mode and coloured by chain with bound metal ions shown as spheres. (B) The chain M is shown as surface mode and coloured by electrostatic potential as calculated using the APBS plugin in PyMOL with red patch indicating negatively-charged residues and blue as positively-charged residues (Baker et al. 2001). (C) Non-FOC interface: dimer interface without metal binding is rich in hydrophobic residues. (D) FOC interface: dimer interface with metal binding in negatively-charged ferroxidase centre (FOC).

Rru-EncFtnSH non-FOC interface				Rru-EncFtnSH FOC interface			
Hydrogen bonds				Hydrogen bonds			
No.	Chain M	Dist. [Å]	Chain N	No.	Chain L	Dist. [Å]	Chain M
1	ARG 22 [HH21]	2.03	ASP 40 [OD2]	1	ASN 21 [HD21]	1.99	THR 47 [OG1]
2	HIS 9 [H]	2.05	GLN 41 [OE1]	2	TYR 39 [HH]	1.66	GLU 32 [OE2]
3	LYS 16 [H]	2.19	ASP 44 [OD2]	3	SER 46 [OG]	3.82	ASN 21 [OD1]
4	TYR 87 [HH]	1.75	GLU 64 [OE2]	4	THR 47 [H]	2.41	ASN 21 [OD1]
5	PHE 89 [H]	2.25	GLU 71 [OE1]	5	HIS 65 [ND1]	3.51	GLU 62 [OE2]
6	THR 90 [H]	2.28	GLU 71 [OE2]	6	ARG 75 [HH11]	2.36	GLU 50 [OE2]
7	LYS 63 [HZ3]	2.18	ASP 77 [OD1]	7	THR 47 [OG1]	1.92	ASN 21 [HD21]
8	ARG 75 [HH21]	2.25	THR 90 [OG1]	8	GLU 32 [OE2]	1.75	TYR 39 [HH]
9	ARG 75 [HH22]	2.18	GLY 92 [O]	9	ASN 21 [OD1]	2.49	THR 47 [H]
10	ARG 75 [HH12]	2.18	GLY 92 [O]	10	GLU 62 [OE2]	3.62	HIS 65 [ND1]
11	ASP 40 [OD2]	1.98	ARG 22 [HH21]				
12	GLN 41 [OE1]	1.86	HIS 9 [H]				
13	ASP 44 [OD2]	2.20	LYS 16 [H]				
14	GLU 64 [OE2]	1.85	TYR 87 [HH]				
15	GLU 71 [OE1]	2.22	PHE 89 [H]				
16	GLU 71 [OE2]	2.35	THR 90 [H]				
17	ASP 77 [OD1]	1.97	LYS 63 [HZ2]				
18	THR 90 [OG1]	2.09	ARG 75 [HH21]				
19	GLY 92 [O]	2.03	ARG 75 [HH12]				
20	GLY 92 [O]	2.28	ARG 75 [HH22]				
Salt bridges				Salt bridges			
No.	Chain M	Dist. [Å]	Chain N	No.	Chain L	Dist. [Å]	Chain M
1	HIS 23 [ND1]	3.43	ASP 37 [OD1]	1	HIS 65 [ND1]	3.72	GLU 62 [OE1]
2	HIS 23 [ND1]	2.63	ASP 37 [OD2]	2	HIS 65 [ND1]	3.51	GLU 62 [OE2]
3	ARG 22 [NH2]	3.94	ASP 40 [OD1]	3	HIS 65 [NE2]	2.70	GLU 61 [OE1]
4	ARG 22 [NE]	3.44	ASP 40 [OD2]	4	ARG 75 [NH1]	2.92	GLU 50 [OE2]
5	ARG 22 [NH2]	2.87	ASP 40 [OD2]	5	ARG 75 [NH1]	3.92	GLU 50 [OE1]
6	LYS 63 [NZ]	3.07	ASP 77 [OD1]	6	GLU 62 [OE1]	3.84	HIS 65 [ND1]
7	LYS 63 [NZ]	3.10	ASP 77 [OD2]	7	GLU 62 [OE2]	3.62	HIS 65 [ND1]
8	ASP 37 [OD1]	3.38	HIS 23 [ND1]	8	GLU 61 [OE1]	2.68	HIS 65 [NE2]
9	ASP 37 [OD2]	2.65	HIS 23 [ND1]				
10	ASP 40 [OD1]	3.86	ARG 22 [NH2]				
11	ASP 40 [OD2]	3.36	ARG 22 [NE]				
12	ASP 40 [OD2]	2.77	ARG 22 [NH2]				
13	ASP 77 [OD1]	2.83	LYS 63 [NZ]				
14	ASP 77 [OD2]	3.12	LYS 63 [NZ]				

Table 3.4.2 Interaction of Rru-EncFtnSH dimer interfaces analysed by PDBePISA.

Three chains of Rru-EncFtnSH with FOC and non-FOC interfaces were calculated by PDBePISA server (Krissinel & Henrick 2007) to estimate the interaction of two types of interfaces. The contribution of metal binding to FOC interface was not listed here, leading to an underestimation of the FOC interface.

3.4.3 Metal/ligand binding in Rru-EncFtnSH

The electron density maps of the initial Rru-EncFtnSH model displayed significant positive peaks in the mFo-DFc map at the centre of the FOC dimer interface (**Figure 3.4.4**). As Rru-EncFtnSH decamerises in the presence of Fe(II), diffraction data were collected at the experimentally determined iron absorption *K*-edge (1.731 Å, 100 K). An anomalous difference Fourier map was calculated using these data (**Figure 3.4.5**).

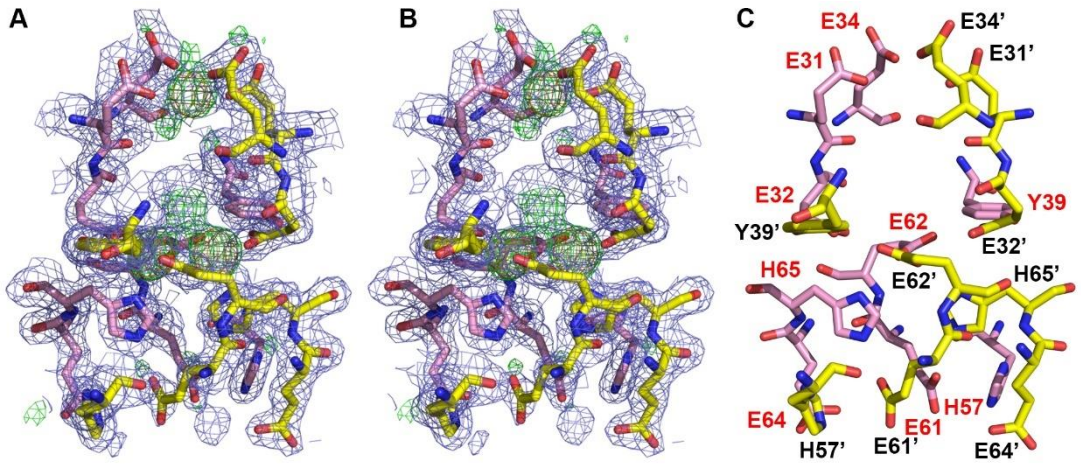


Figure 3.4.4 Putative metal/ligand-binding sites in Rru-EncFtnSH FOC dimer interface.

A dimer of Rru-EncFtnSH is shown as sticks and coloured by pink and yellow. (A–B) Wall-eyed stereo view of metal/ligand-binding sites in Rru-EncFtnSH FOC dimer interface with 2mFo-DFc electron density superimposed, shown in blue mesh and contoured at 1.0 σ. The mFo-DFc map is shown in green mesh and contoured at 4 σ. The NCS averaged anomalous map shown in orange mesh and contoured at 20 σ. (C) Metal/ligand binding sites labelled in red are from chain M in pink while black labels with a prime symbol correspond to chain L in yellow.

The anomalous difference map showed two peaks between residues E32, E62, H65 and Y39 of two adjacent monomers with an NCS averaged anomalous peak height around 67 σ, and a statistically smaller peak near residues E31, E34 of two neighbouring chains with an NCS averaged anomalous peak height of around 45 σ (**Figure 3.4.4**). Modeling metal ions into these peaks and refinement of the anomalous scattering parameters allowed us to identify these as two iron ions and a calcium ion, respectively. When iron ions were modelled in the E31-E34 positions, the B-factors and occupancies failed to be refined to a reasonable range. While ICP-MS data indicated that there were negligible amounts of calcium in the purified protein (**Table 3.3.1**), the presence of 140 mM calcium acetate in the crystallisation solutions favoured the coordination of calcium at these sites. It is worth mentioning that the E31

and E34 residues might have also coordinated iron or calcium in the crystal structure, leading to an averaged smaller anomalous peak. An additional region of asymmetric electron density near the di-iron binding site in the mFo-DFc map was modeled as glycolic acid (GOA), a breakdown product of the PEG 3350 used for crystallisation (**Figure 3.4.5**).

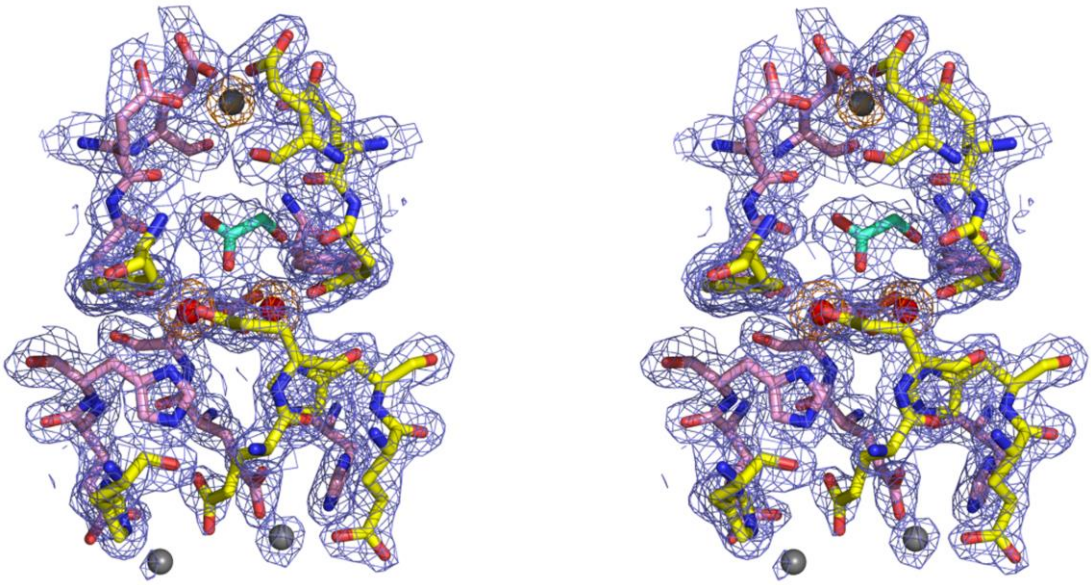


Figure 3.4.5 Wall-eyed stereo view of the FOC dimer interface of Rru-EncFtnSH.

Protein residues are shown as sticks with pink and yellow carbons for the different subunits. Iron ions are shown as red spheres and calcium ions as grey spheres; and the glycolic acid (GOA) is shown with green carbon atoms bridging the di-iron. The 2mFo-DFc electron density map is shown as a blue mesh contoured at 1.0 σ. The NCS averaged anomalous map obtained from data collected at the Fe *K*-edge is superimposed as an orange mesh and contoured at 20 σ.

3.4.4 Ferroxidase centre (FOC)

In the FOC dimer interface of Rru-EncFtnSH there is a two-fold symmetrical di-iron centre, with two iron ions coordinated by E32, E62, H65 and Y39 from two adjacent chains. Each iron is coordinated by one histidine nitrogen (H65) and five carboxyl oxygens. Four of these oxygen ligands are supplied by bidentate glutamate carboxyl (E32' from chain L and E62 from chain M). The other one arises from monodentate E62' of chain L. E62 from both chains straddles two iron ions.

The Fe–Fe distance is 3.5–3.7 Å, Fe–Glu–O distances between 2.3 and 2.5 Å, and Fe–His–N distances of 2.5 Å (**Figure 3.4.6**). This coordination geometry is consistent with

the di-nuclear ferroxidase center found in ferritin (Bertini et al. 2012). The distance of Fe–Tyr39–O is 4.8 Å, which is beyond the classic iron coordination distance (Harding et al. 2010). Tyr39 is highly conserved in ferritin family proteins; and is suggested to be involved in facilitating oxidation by acting as a single-electron molecular capacitor (Ebrahimi, Hagedoorn & Wilfred R Hagen 2013).

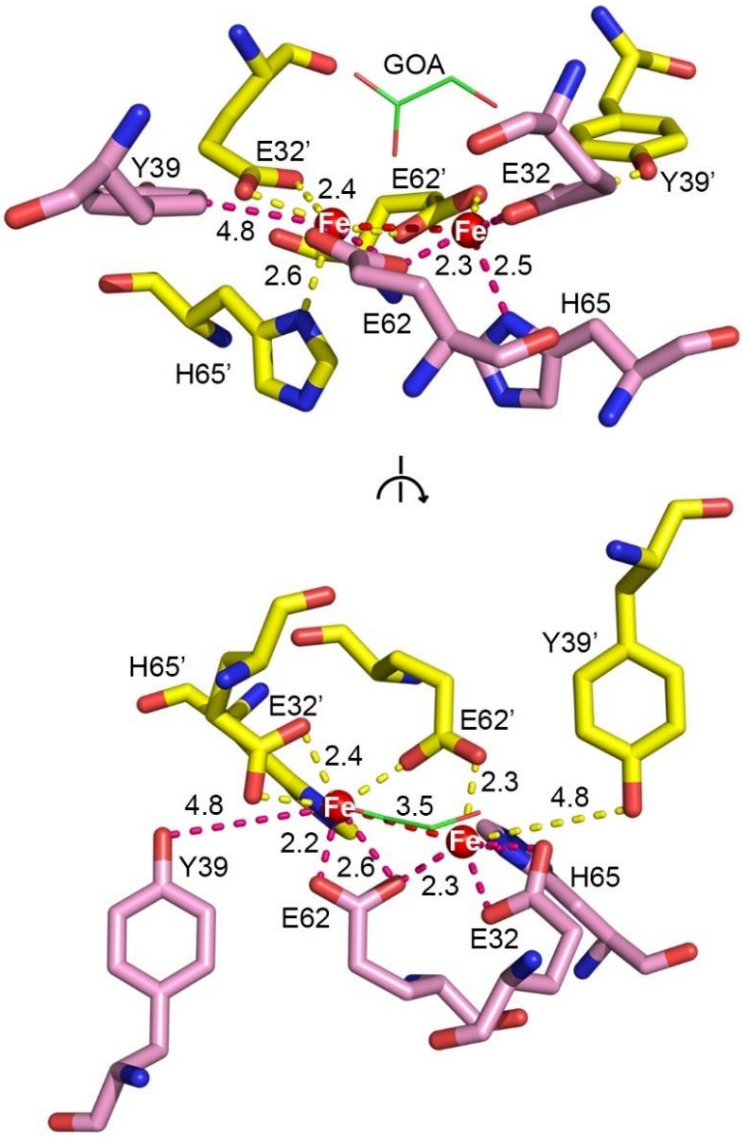


Figure 3.4.6 Ferroxidase centre (FOC) of Rru-EncFtnSH.

The FOC consists of E32, E62, H65A and Y39 from two adjacent monomers shown in pink (chain M) and yellow (chain L) sticks. Glycolic acid (GOA) is highlighted in green carbons bound above the di-iron (red spheres). Distances between iron and amino acid residues or the second iron are labelled as numbers with Å as the unit.

Due to experimental limitations it was not possible to differentiate the oxidation state of the iron in the FOC. X-ray absorption spectroscopy can be used to study the dynamic iron oxidation states in protein crystals (Antonyuk & Hough 2011) and solutions (Toussaint et al. 2009); and Mössbauer spectroscopy and electron paramagnetic resonance (EPR) (Ebrahimi et al. 2012; Bou-Abdallah et al. 2014) can be applied to examine the valence state of iron in protein solutions.

It is interesting to note that both monomer and decamer fractions of Rru-EncFtnSH were able to crystallise into an annular decamer structure without any additional iron ions in the crystallisation or soaking conditions, indicating that Rru-EncFtnSH scavenged even trace amount of iron ions from environment. In our experiments it was impossible to get iron-free structures even when iron-free monomer fractions were applied. The absence of iron supplementation during protein production or crystallisation still gave rise to a fully occupied FOC with di-iron in the final structure, implying that the FOC has a very high affinity for iron ions.

3.4.5 Calcium binding sites

Apart from iron ions bound in the Rru-EncFtnSH FOC, there are a number of calcium ions bound by surface residues (D44, D48, E50, W72, N76 and E91) as no anomalous peaks around these surface residues have been detected. The calcium ions coordinated in the FOC dimer interface of Rru-EncFtnSH can be categorised into two sites: metal ion entry site and metal ion exit site (**Figure 3.4.7**).

The metal ion entry site locates at the inner circumference of the decamer structure with one Ca(II) per dimer. It is composed of residues E31 and E34 from two chains. Ca(II) is bound by four carboxyl oxygens from monodentate E31 and E34 of two chains, along with three water molecules. The coordination distances between Ca(II) and binding residues, or water are 2.5 ± 0.2 Å. This heptacoordinate geometry is common in crystal structures with calcium ions bound (**Figure 3.4.7 A**) (Katz et al. 1996). The metal ion exit sites are partially occupied by Ca(II) present at the outer circumference of the decamer structure. In **Figure 3.4.7 B**, two Ca(II) are bound by H57, E61 and E64 from both chains in the FOC dimer interface and are 5.2 Å apart. Fe–Glu–O distances are between 2.6 and 3.7 Å and the Fe–His–N distances are 3.5

and 4.7 Å. It has not escaped our notice that in the asymmetric unit Ca(II) is missing in the metal ion exit site of eleven dimers while another two dimers only contain a single Ca(II).

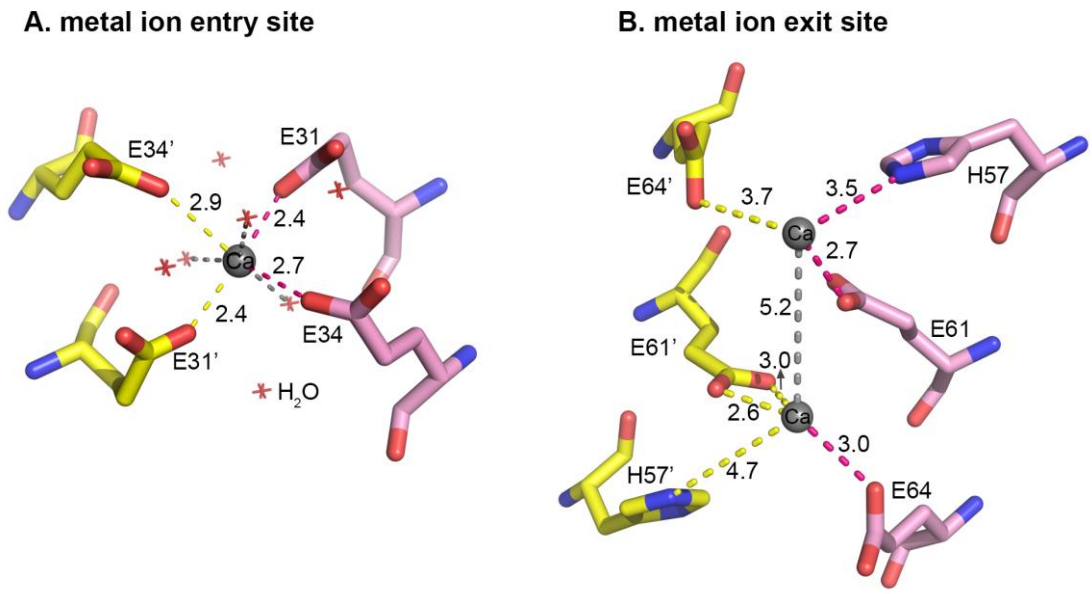


Figure 3.4.7 Calcium binding sites in Rru-EncFtnSH FOC dimer interface.

Ca(II) is shown in grey spheres coordinated by two monomers shown in sticks and coloured in yellow and pink. Distances between Ca(II) and binding residues or Ca–Ca are labelled as numbers with Å as the unit. (A) Metal ion entry site consists of E31 and E34 from two adjacent monomers. Ca(II) is coordinated by three water molecules (red crosses) with grey dashed lines linked as well as carboxylate oxygens from E31 and E34 of two chains. (B) Metal ion exit site consists of H57, E61 and E64 from both yellow and pink chains coordinating two Ca(II).

As the FOC is located between the two calcium binding sites, iron has to translocate from either metal ion entry site or exit site to the FOC. This implies that the calcium binding sites in the crystal structure are very likely to be an alternative iron transition site under native conditions.

To validate the iron binding in metal ion entry/exit sites, efforts have been made to increase the anomalous peaks of iron in calcium binding sites by soaking Rru-EncFtnSH crystals aerobically in cryo-protectant solutions with 1 mM acidic $Fe(NH_4)_2(SO_4)_2$ from 2 minutes to 1 hour, and then flash soaking back to cryo-protectant solutions to remove free iron ions, or without back soaking. However, they all failed to increase the occupancy of these sites by iron as would be indicated by higher peaks in the anomalous Fourier map arising iron in metal ion entry site. This

might be related to the low concentration of iron in aerobic soaking, leading to about 1.3 iron per monomeric Rru-EncFtnSH, which is insufficient to occupy the metal ion entry site, or compete with 140 times more calcium from the crystallisation solution. We also observed that with longer iron soak under aerobic conditions, diffraction from the crystals decayed rapidly, indicating higher levels of radiation damage from the absorbance of X-rays by the additional iron ions in the crystal.

Rru-EncFtnSH was shown to decamerise in solution in the presence of Co(II), Zn(II) or Fe(II) (**Figure 3.5.2** and **Figure 3.5.4**). Co(II) resembles Fe(II) in terms of binding geometry and distance in protein structures (Harding et al. 2010) and its ability to form dinuclear centres. Attempts have been made to obtain an Rru-EncFtnSH structure with Co(II) coordinated instead of Fe(II) by co-crystallising the iron-free monomer fractions with $CoCl_2$ or crystal soaking in $CoCl_2$. However, these experiments failed to yield a structure with no iron ions bound; or the diffraction resolution was not high enough to determine the metal binding in structures. Apart from Co(II), Rru-EncFtnSH crystals have also been soaked in Zn(II), of which the metal binding sites were still occupied by iron and calcium ions. Collectively, all those observations indicate that Rru-EncFtnSH has higher binding affinity towards Fe(II) than any metal of the ions tested in our protein assembly experiments (**Figure 3.5.2** and **Figure 3.5.4**).

3.4.6 Unstructured monomeric Rru-EncFtnSH determined by ion mobility mass spectrometry

Since the monomer structure of Rru-EncFtnSH was not available from X-ray crystallography, ion mobility-mass spectrometry (IM-MS) was used to analyse the monomer conformation in the gas phase, based on different mobilities in an electric field. The IM-MS analysis of the monomeric Rru-EncFtnSH at pH 8.0 revealed that the metal-free protein monomer exists in a wide range of charge states (+6 to +16) and adopts many conformations in the gas phase with CCSs ranging from 12 nm^2 to 26 nm^2 (**Figure 3.4.8**). These observations are indicative of an unstructured protein with little secondary or tertiary structure (Beveridge et al. 2014). Thus, IM-MS studies indicate that the monomer is at least partially unstructured and that the higher order structure is mediated/stabilized by metal binding, an observation that is in agreement with our solution studies. Taken together, these results suggest that di-iron binding,

forming the FOC in Rru-EncFtnSH, is required to stabilise the 4-helix bundle dimer interface, essentially reconstructing the classical ferritin-like fold. Once stabilised, these dimers readily associate as pentamers, and the overall assembly adopts the decameric ring arrangement observed in the crystal structure.

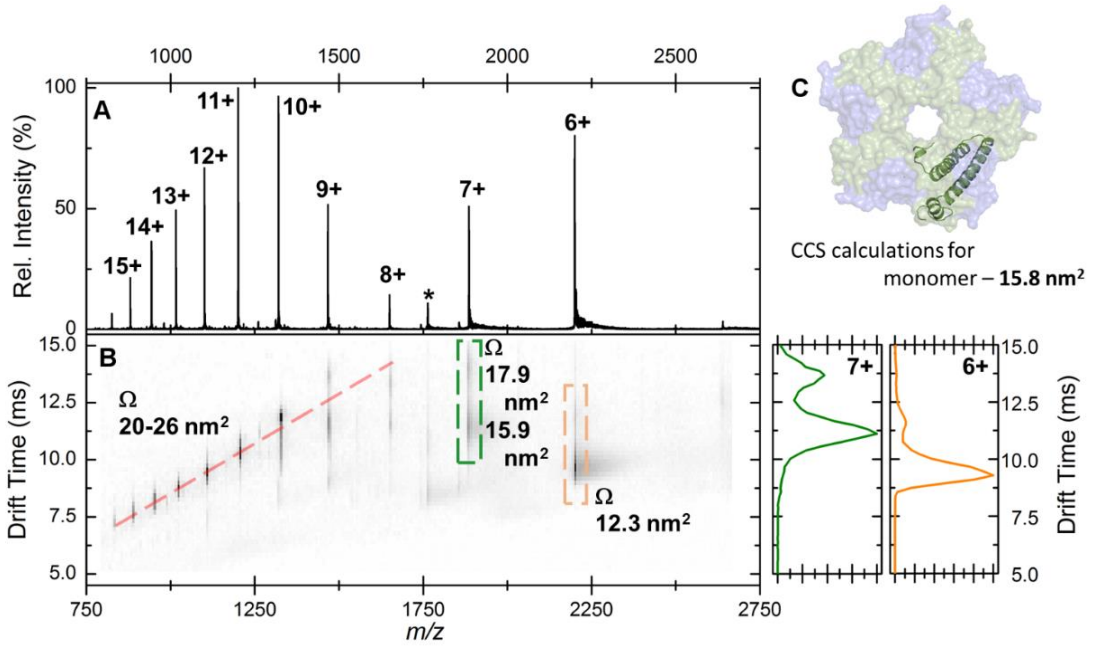


Figure 3.4.8 Native IM-MS analysis of the apo-Rru-EncFtnSH monomer.

(A) Mass spectrum of apo-Rru-EncFtnSH monomer acquired from 100 mM ammonium acetate pH 8.0 under native MS conditions. The charge state distribution observed is bimodal, with peaks corresponding to the 6+ to 15+ charge states of apo-monomer Rru-EncFtnSH (neutral average mass 13194.3 Da). (B) left, the arrival time distributions (ion mobility data) of all ions in the apo-Rru-EncFtnSH charge state distribution displayed as a greyscale heat map (linear intensity scale). (B) right, the arrival time distribution of the 6+ (orange) and 7+ (green) charge state (dashed coloured-box) has been extracted and plotted; the arrival time distributions for these ion is shown (ms), along with the calibrated collision cross section, Ω (nm^2). (C) The collision cross section of a single monomer unit from the crystal structure of the Fe-loaded Rru-EncFtnSH decamer was calculated to be 15.8 nm^2 using IMPACT v. 0.9.1. The +8 to +15 protein charge states have observed CCS between 20–26 nm^2, which is significantly higher than the calculated CCS for a Rru-EncFtnSH monomer taken from the decameric assembly crystal structure (15.8 nm^2). The mobility of the +7 charge state displays broad drift-time distribution with maximal consistency with CCS of 15.9 and 17.9 nm^2. Finally, the +6 charge state of Rru-EncFtnSH has mobility consistent with a CCS of 12.3 nm^2, indicating a more compact / collapsed structure. It is clear from this data that apo-Rru-EncFtnSH exists in several gas phase conformations. The range of charge states occupied by the protein (6+ to 15+) and the range of CCS in which the protein is observed (12.3 nm^2–26 nm^2) are both large. In addition, many of the charge states observed have higher charge than the theoretical maximal charge on spherical globular protein, as determined by the De La Mora relationship ($ZR = 0.0778\sqrt{m}$; for the Rru-EncFtnSH monomer ZR = 8.9) (Fernandez de la Mora 2000). As described by Beveridge et al., all these factors are indicative of a disordered protein (Beveridge et al. 2014).

3.4.7 Comparison of Rru-EncFtn and a bacterial ferritin

Rru-EncFtn, as a representative of EncFtn, was compared with a classical ferritin PfFtn—a bacterial ferritin from *P. furiosus* (KEGG ID: PF0742, PDB ID: 2JD7) in four aspects: secondary structure, tertiary structure, quaternary structure, as well as metal binding sites.

Although the primary sequence of Rru-EncFtn was only 17% shorter than PfFtn (Tatur et al. 2007), an FOC dimer of Rru-EncFtn is equivalent to a monomer of PfFtn as a four-helical bundle of ferritin fold (**Figure 3.4.9 A/B**). The linker connecting α2 and α3 of PfFtn was congruent with the start of the C-terminal helix of one Rru-EncFtn monomer and the N-terminal 3_{10} helix of the second monomer. The structural alignment of the Rru-EncFtnSH FOC dimer with the monomeric PfFtn also revealed a striking similarity, indicating that the two families of proteins have essentially the same architecture around the di-iron center (**Figure 3.4.9 B**). When superimposing the biological assembly of Rru-EncFtnSH with PfFtn, the central ring of Rru-EncFtnSH corresponded to the external surface of PfFtn, while the outer circumference of Rru-EncFtnSH was congruent with the inner mineralisation surface of PfFtn (**Figure 3.4.9 C**). The size of the Rru-EncFtn decamer is 7 nm in contrast with 12 nm for the PfFtn 24-mer. The alignment of decameric Rru-EncFtnSH and 24-meric PfFtn suggested that the FOC of Rru-EncFtnSH is essentially oriented the opposite way in terms of the quaternary arrangement in the ring structure.

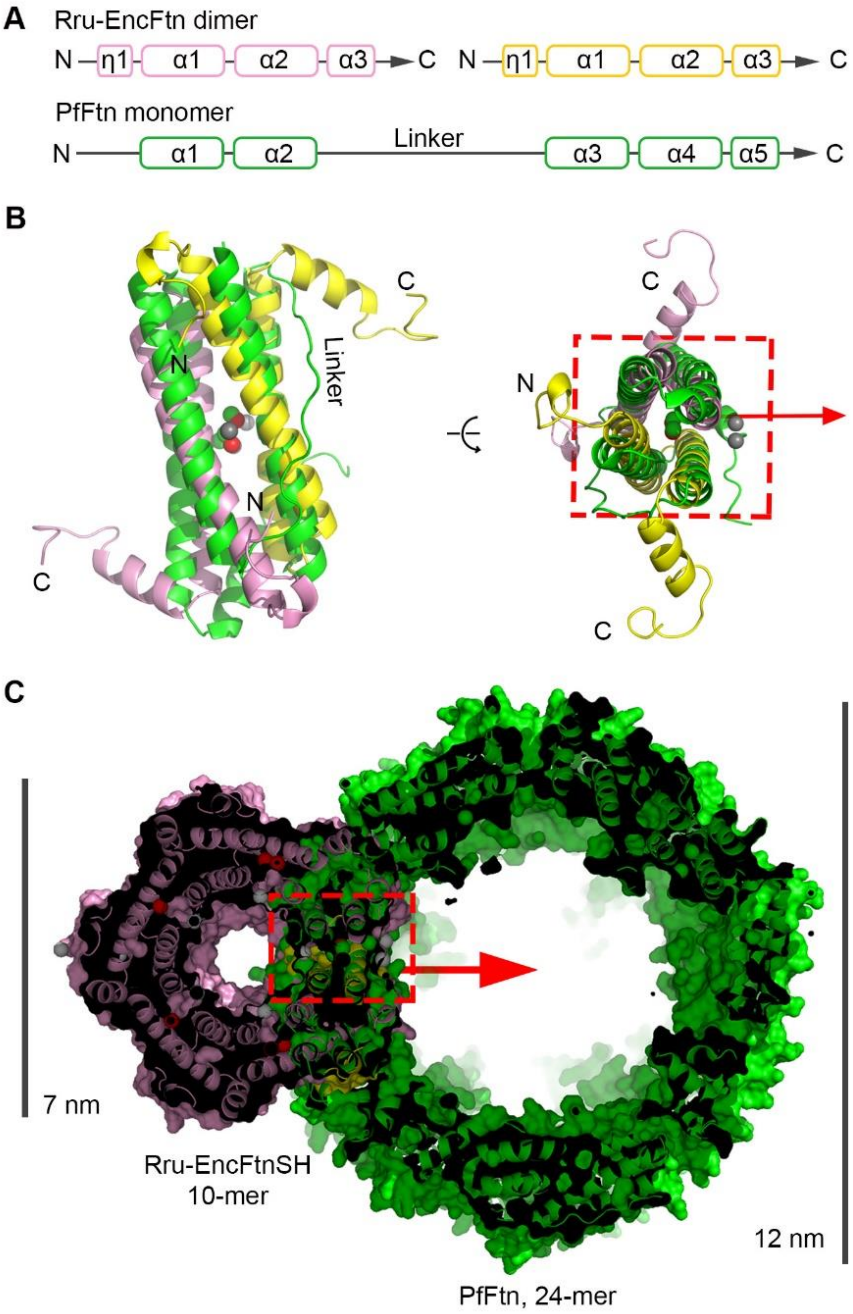


Figure 3.4.9 Structure alignments of Rru-EncFtn and *P. furiosus* bacterial ferritin (PfFtn).

(A) Secondary structure overview of a dimer of Rru-EncFtn and a PfFtn (KEGG ID: PF0742, PDB ID: 2JD7) monomer (B) Structural alignment of the Rru-EncFtnSH FOC dimer (pink and yellow alpha helices in cartoon view; two Fe(III) shown in red spheres; a Ca(II) shown in a white sphere) and a PfFtn monomer (green alpha helices in cartoon view; three Fe(III) shown in green spheres). The proteins align with a $RMSD_{C\alpha}$ of 2.52 Å over α1-α2 residues. The red arrow and dashed rectangle highlight the top view of alignment and indicate the proposed iron moving route from the ferroxidase centre to the mineralisation site in the ferritin cage. (C) Biological assembly alignment between Rru-EncFtnSH and PfFtn shown in surface and cartoon modes. The red dashed rectangle with the red arrow indicates the top view in (B). The scale bar indicates the diameter of Rru-EncFtnSH decamer as 7.6 nm and PfFtn 24-mer as 12 nm.

3.4.8 Alignment of the metal binding sites of Rru-EncFtnSH and a bacterial ferritin

The metal binding site alignment of Rru-EncFtn and PfFtn has been highlighted in **Figure 3.4.10**. The superimposing FOCs of Rru-EncFtnSH and PfFtn presented the fact that with respect to the quaternary structure of the proteins, the FOC of Rru-EncFtn faces in the opposite direction relative to the bacterial ferritin, and is essentially inside out regarding the iron storage space (**Figure 3.4.9 B and Figure 3.4.10 A**). The FOC of PfFtn coordinates three iron ions: with Fe_A and Fe_B bound in the conserved di-iron centre; with a third iron ion, Fe_C, bound in site C, at 7.5 and 6.3 Å distance to Fe_A and Fe_B, respectively (Tatur et al. 2007). Site C was suggested to be a weaker binding site, a conserved gateway for Fe(II) translocation to the FOC, and likely to be the exit site of Fe(III) from the FOC as well (Ebrahimi et al. 2012).

The alignment of the FOCs of Rru-EncFtnSH and PfFtn revealed the conservation of the di-iron centre between these ferritin family proteins. The main difference between the two proteins lies in the totally symmetric FOC of Rru-EncFtnSH, which is a consequence of its formation at the interface of two monomers; and the asymmetric FOC of PfFtn, at the centre of the single-chain four-helix bundle. In the PfFtn residues equivalent to H65 of one subunit of Rru-EncFtnSH in the FOC dimer correspond to E130 in the FOC of PfFtn, and one of the tyrosine residues (Y39) in the Rru-EncFtnSH FOC dimer is absent in the PfFtn (**Figure 3.4.10 B**). The different coordination environments of Fe_A and Fe_B in PfFtn gives rise to a stronger binding affinity for iron in the Fe_A site and weaker binding affinities in the Fe_B and Fe_C sites (Ebrahimi et al. 2012). It is postulated that the two iron ions bound in the FOC of Rru-EncFtnSH have equivalent binding energies due to the symmetrical FOC. The alignment of metal binding sites of two proteins also uncovered a distinct metal binding site in Rru-EncFtnSH, which bound a calcium ion in the crystal structure (**Figure 3.4.10 A, left side**). The presence of the novel calcium binding site in Rru-EncFtnSH suggested a distinct metal ion entry route from classical ferritins. The overlay of the metal ion exit site in Rru-EncFtnSH and site C of PfFtn displayed a poor correspondence between the two proteins, except for the E61' of Rru-EncFtnSH, which superimposed on E49 of PfFtn (**Figure 3.4.10 C**). This showed the relatively low conservation of the third iron binding site of PfFtn and metal ion exit site of Rru-EncFtnSH. Taken together,

the open structure of Rru-EncFtnSH along with the everted iron storage directionality led to the question: how can iron be stored in Rru-EncFtn while iron ions move from the inner channel to the outer surface? Is Rru-EncFtnSH catalytically active to catalyse the Fe(II) oxidation?

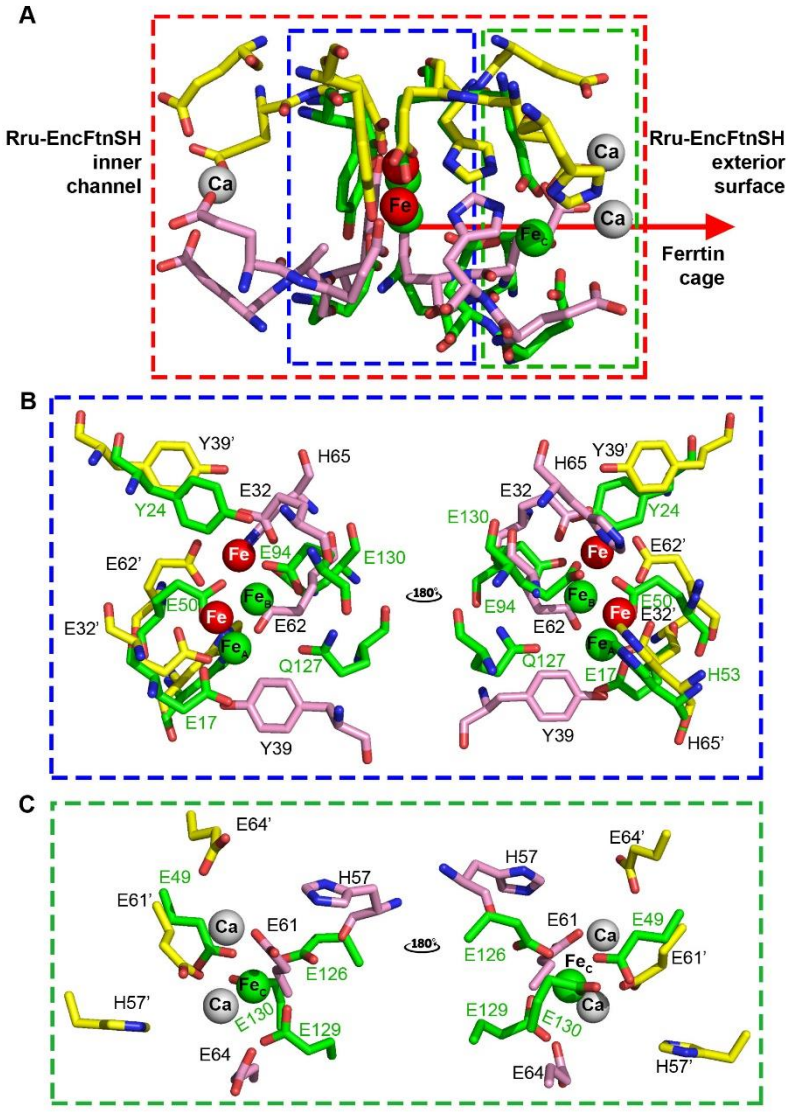


Figure 3.4.10 Superimposition of metal binding sites in the Rru-EncFtnSH dimer and the PfFtn monomer.

(A) Metal binding sites shown in stick mode with Rru-EncFtnSH displayed as yellow and pink sticks; with the PfFtn monomer shown in green. The red arrow and red dashed rectangle are consistent with panels B and C in **Figure 3.4.9**, highlighting the iron storage direction of PfFtn. The blue dashed rectangle highlights the overlay of ferroxidase centres. The green dashed rectangle highlights the comparison of non-FOC binding sites. (B) Side view of blue dashed area from (A) with Rru-EncFtnSH residues labelled in black and PfFtn residues labelled in green. Red spheres highlight the two iron ions coordinated by the Rru-EncFtnSH FOC including E32, E62, H65 and Y39 from pink chain and E32', E62', H65' and Y39' from yellow chain. Green spheres highlight Fe_A coordinated by E17, E50 and H53 of PfFtn; with Fe_B coordinated by E50, E94, E130. Q127 coordinates both Fe_A and Fe_B at 3.7 Å (Tatur et al. 2007). (C) Side view of green dashed area from (A). White spheres highlight two Ca(II) coordinated in the Rru-EncFtnSH dimer interface by H57, E61 and E64 from the pink chain and H57', E61' and E64' from yellow chain. The green sphere highlights the third iron ion (Fe_C) binding site E49, E126, E129 and E130 in PfFtn, among which only E49 is conserved between PfFtn and Rru-EncFtnSH (E61'). (A) and (B) suggest the conservation of ferroxidase centre between Rru-EncFtnSH and PfFtn. (A) also delineates a Ca(II) bound at the entry site (E31, E34, E31' and E34') in Rru-EncFtnSH, which does not show equivalent residues in PfFtn. (C) indicates the variability of non-FOC metal binding sites.

3.5 Assembly and disassembly of Rru-EncFtnSH

3.5.1 Fe(II)-dependent Rru-EncFtnSH decamerisation in gas phase

To quantify the Rru-EncFtnSH oligomerisation in the presence of Fe(II) *in vitro*, nano-electrospray ionisation mass spectrometry (nESI-MS, or native MS) was performed by Sam Huges and Dr. David Clarke (School of Chemistry, The University of Edinburgh). 50 µM Rru-EncFtnSH monomer fractions in 100 mM NH_4 Acetate buffer (pH 8.0) were mixed with $FeCl_2$ at various concentrations prior to nESI-MS analysis. nESI-MS allows the measurement of the molecular weight of native proteins without disturbing protein assembly (Konijnenberg et al. 2013).

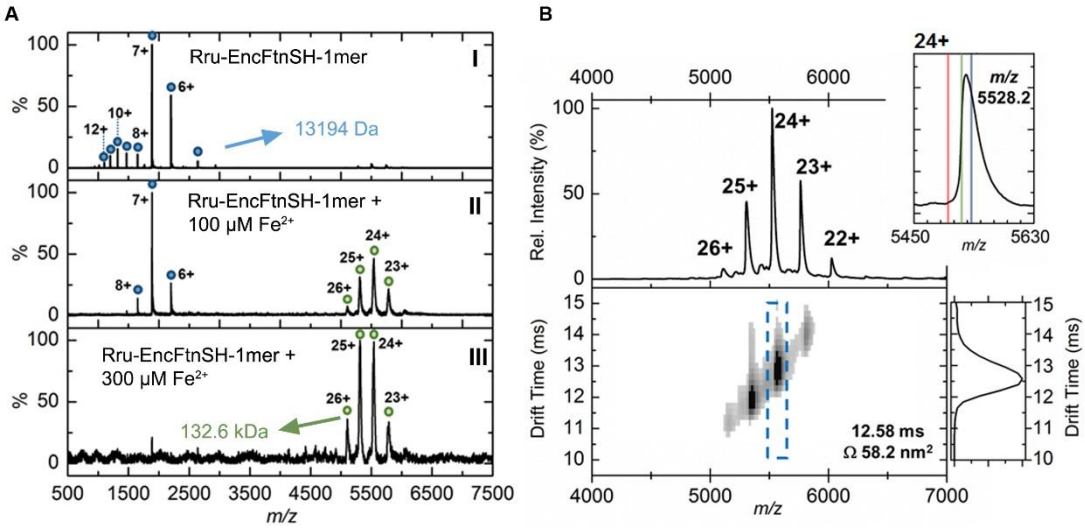


Figure 3.5.1 Native mass spectrometry and ion mobility analysis of iron loading in Rru-EncFtnSH.

All spectra were acquired in 100 mM ammonium acetate, pH 8.0 with a protein concentration of 50 µM. (A) Native nanoelectrospray ionization (nESI) mass spectrometry of Rru-EncFtnSH at varying iron concentrations. A-I, nESI spectrum of iron-free Rru-EncFtnSH displays a charge state distribution consistent with Rru-EncFtnSH monomer (blue circles, 13,194 Da). Addition of 100 µM (A-II) and 300 µM (A-III) Fe(II) results in the appearance of a second higher molecular weight charge state distribution consistent with a decameric assembly of Rru-EncFtnSH (green circles, 132.6 kDa). (B) Ion mobility (IM)-MS of the iron-bound holo-Rru-EncFtnSH decamer. Top, Peaks corresponding to the +22 to +26 charge states of a homo-decameric assembly of Rru-EncFtnSH are observed (132.6 kDa). Top Insert, Analysis of the 24+ charge state of the assembly at m/z 5528.2 Th. The theoretical average m/z of the 24+ charge state with no addition metals bound is marked by a red line (5498.7 Th); the observed m/z of the 24+ charge state indicates that the Rru-EncFtnSH assembly binds between 10 (green line, 5521.1 Th) and 15 Fe ions (blue line, 5532.4 Th) per decamer. Bottom, the arrival time distributions (ion mobility data) of all ions in the Rru-EncFtnSH charge state distribution displayed as a greyscale heat map (linear intensity scale). Bottom right, the arrival time distribution of the 24+ charge state (dashed blue box) has been extracted and plotted. The drift time for this ion is shown (ms), along with the calibrated collision cross section (CCS), Ω (nm^2).

In **Figure 3.5.1 A**, the mass to charge ratio of the monomeric Rru-EncFtnSH are plotted, showing the charge state distribution of the protein sample. The charge states from 12+ to 6+ correspond to 13.194 kDa, consistent with theoretical monomer size of 13.19453 kDa. Under these conditions, no significant higher order assembly was observed in the monomeric Rru-EncFtnSH and the protein did not have any coordinated metal ions. Titration with Fe(II) directly before native MS analysis resulted in the appearance of a new charge state distribution, consistent with an Rru-EncFtnSH decameric assembly (+22 to +26; 132.65 kDa) (**Figure 3.5.1 A II–III**). After instrument optimisation, the mass resolving power achieved was sufficient to assign iron-loading in the complex to 10–15 iron ions per decamer (**Figure 3.5.1 B**, inset top right), consistent with the presence of 10 iron ions in the FOC and the coordination of iron in the E31/34-site (metal ion entry site) occupied by calcium in the crystal structure (Δmass ~0.67 kDa). MS analysis of Rru-EncFtnSH after addition of further Fe(II) did not result in iron loading higher than this stoichiometry. These data suggest that the decameric assembly of Rru-EncFtnSH does not accrue iron in the same manner as classical ferritin, which is able to sequester more than 4000 iron ions within its nanocage (Mann et al. 1986).

Ion mobility-mass spectrometry (IM-MS), using a Travelling Wave Ion Mobility (Triwave) system from Waters Synapt instruments, separates molecules based on size and shape in addition to the measurement of mass to charge ratios. IM-MS characterises a molecular structure by estimating the collision cross section values (CCS) (Kanu et al. 2008). IM-MS analysis of the Rru-EncFtnSH decameric assembly, collected with minimal collisional activation, suggested that it consists of a single conformation with a CCS of 58.2 nm^2 (**Figure 3.5.1 B**). The CCS can also be calculated as 58.7 nm^2 using the 'IMPACT' tool based on the crystal structure of Rru-EncFtnSH (Marklund et al. 2015), which is in agreement with the IM-MS result.

Rru-EncFtnSH bound up to fifteen iron ions per decamer as suggested by native mass spectrometry, whereas the crystal structure displayed ten iron ions in the five FOCs, five Ca(II) in metal entry site and a number of Ca(II) bound by surface residues. The discrepancy between this iron loading experiment and the MS and crystallographic studies suggests that the protein has a number of weak surface metal binding sites; and

a subset of these may be those that are occupied by the calcium ions in the crystal structure.

3.5.2 Metal-dependent assembly of Rru-EncFtnSH

Since metalloproteins are able to bind metal ions in various affinities as suggested by the Irving-William series (Waldron & Robinson 2009), the possibility that the assembly of Rru-EncFtnSH is influenced by various metals was investigated. Native PAGE and analytical gel-filtration were used to evaluate protein oligomerisation in the presence of divalent metal ions. To further differentiate the protein preference for ferric or ferrous iron, solutions of iron salts in the different oxidation states were employed in these experiments.

3.5.2.1 Native PAGE

Rru-EncFtnSH monomer and decamer fractions were separately pooled from MM purification and titrated with equimolar concentrations of metal ions for 2–3 hrs before being subjected to native PAGE (**Figure 3.5.2 A–B**) and SDS-PAGE analysis (**Figure 3.5.2 C–D**). Native PAGE separates proteins based on charge to mass ratio. Compared to the monomer sample without metal ions, bands of higher mass-charge ratio were visible on the native PAGE gel in the presence of Fe(II), Co(II), Zn(II) or Mn(II) at 1:1 molar ratio (**Figure 3.5.2 A**), of which only the titration of Fe(II) led to stable decamer and dimer bands when subjected to SDS-PAGE (**Figure 3.5.2 C**). By contrast, the effect of Ca(II), Mg(II) or Fe(III) supplementation on Rru-EncFtnSH decamerisation was minimal at equimolar ratio (**Figure 3.5.2 A**). The addition of metal to the Rru-EncFtnSH decamer did not show significant changes in the assembly by native PAGE (**Figure 3.5.2 B**). There was a faint band around 60 kDa on SDS-PAGE (**Figure 3.5.2 D**) when Ca(II) was loaded to equal molar Rru-EncFtnSH decamer, suggesting that Ca(II) could contribute to its stability of Rru-EncFtnSH to a certain degree (**Figure 3.5.2 D**). It also confirms the Ca(II) binding in metal ion entry site of Rru-EncFtnSH structure.

These results revealed that Rru-EncFtnSH binds to Fe(II), Co(II), Zn(II) and Mn(II) at equimolar concentrations to produce species consistent with the formation of a decamer. While Ca(II), Mg(II) and Fe(III) do not induce the formation of higher order

complexes of the monomer; however, these results do not eliminate the possibility that the multimerisation of Rru-EncFtnSH can be induced by higher concentrations of these metals.

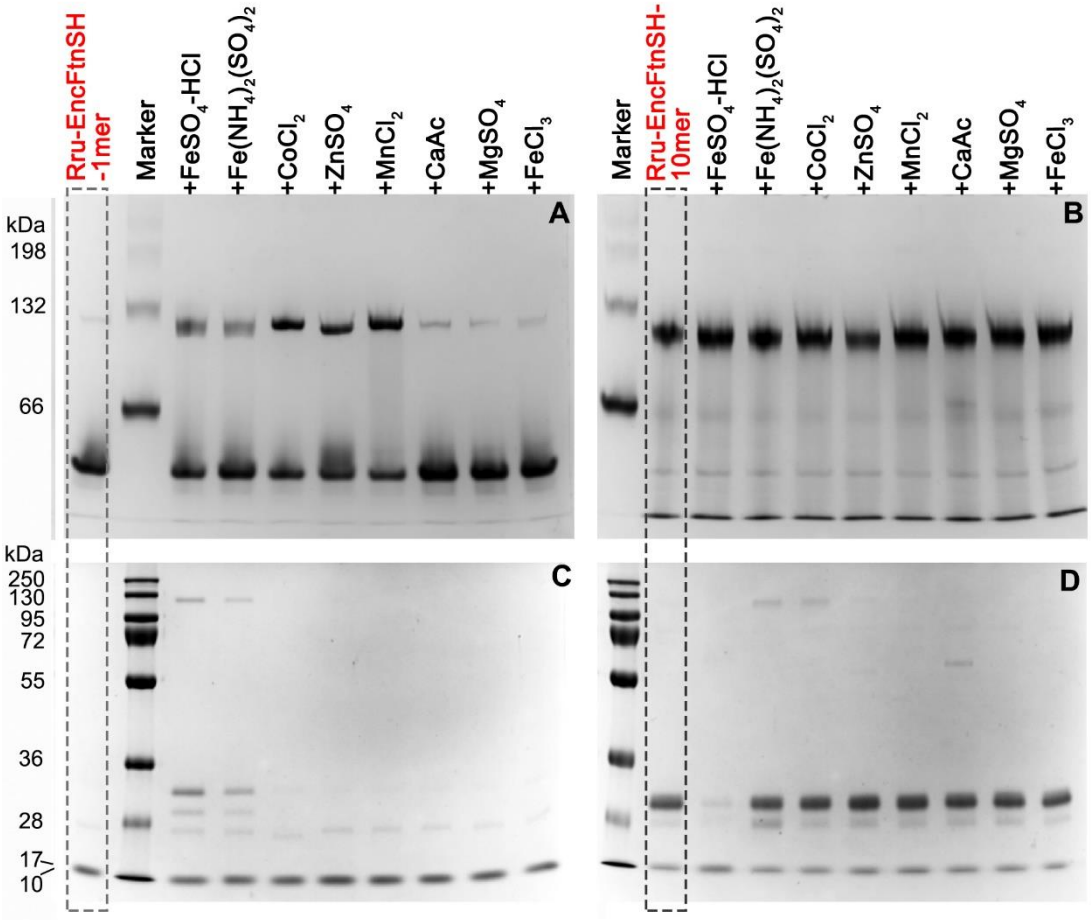


Figure 3.5.2 PAGE analysis of the effect of metal ions on the oligomeric state of Rru-EncFtnSH.

50 µM Rru-EncFtnSH monomer fractions (Rru-EncFtnSH-1mer) or decamer (Rru-EncFtnSH-10mer) samples were mixed with equal molar metal ions including Fe(II), Co(II), Zn(II), Mn(II), Ca(II), Mg(II) and Fe(III), which were analysed by Native PAGE alongside SDS-PAGE. (A) 10% Native PAGE analysis of Rru-EncFtnSH monomer fractions mixed with various metal solutions; (B) 10% Native PAGE analysis of Rru-EncFtnSH decamer fractions mixed with various metal solutions; (C) 15% SDS-PAGE analysis on the mixtures of Rru-EncFtnSH monomer fractions and metal solutions; (D) 15% SDS-PAGE analysis on the mixtures of Rru-EncFtnSH decamer fractions and metal solutions.

3.5.2.2 Analytical Gel-filtration

In order to validate the influence of metal binding on protein assembly as indicated by Native PAGE and SDS-PAGE, the metal titration experiment was performed using Rru-EncFtnSH monomer fractions at 90 µM with subsequent analysis by analytical gel-filtration (AGF) chromatography using Superdex 200 PC 3.2/30 column (**Figure 3.5.4**). To assess the possibility that protein multimerisation was affected by the protein concentration AGF was performed at 90 and 700 µM protein concentration (**Figure 3.5.3**). The slight shift in the main chromatographic peak is consistent with the dimerisation of Rru-EncFtnSH, however the relative proportions of the two peaks seen do not change. This indicates that while the monomer/dimer equilibrium is influenced by protein concentration the formation of Rru-EncFtnSH decamer is not influence by protein concentration.

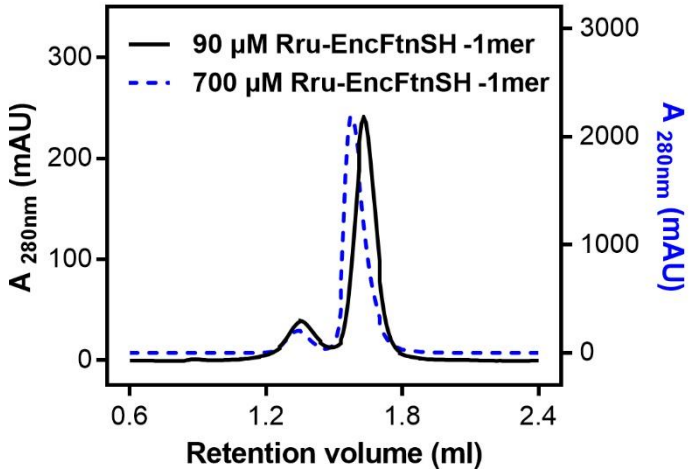


Figure 3.5.3 Concentration-dependent Rru-EncFtnSH assembly analysed by AGF using Superdex 200 PC 3.2/30 column.

700 µM Rru-EncFtnSH monomer fractions had been diluted into 90 µM in GF buffer; both samples were resolved by AGF with the left y-axis demonstrating the 90 µM protein while the right y-axis highlighting the 700 µM protein. The chromatogram comparison shows that Rru-EncFtnSH monomer peak shifted towards dimer at the higher concentration, suggesting the concentration-dependent protein equilibrium between dimer and monomer states.

This 'metal loading-AGF' experiment was performed over three consecutive days, so the protein monomer was run each day as a control to demonstrate the influence of time-dependent protein oligomerisation. The AGF results showed that Rru-EncFtnSH tends to equilibrate between monomer and decamer; and forms higher-order aggregates over time (**Figure 3.5.4**). The addition of equimolar concentrations of

Fe(II), Co(II) or Zn(II) to monomeric Rru-EncFtnSH induces the formation of protein decamers; while the addition of Fe(III), Ca(II) and Mg(II) has little effect on protein assembly. However, AGF results disagreed with native PAGE data with respect to the minimal impact of Mn(II) on Rru-EncFtnSH assembly. More experiments need to be done to confirm the Mn(II) effect on the protein assembly as well as how metal to protein loading ratios alter the assembly.

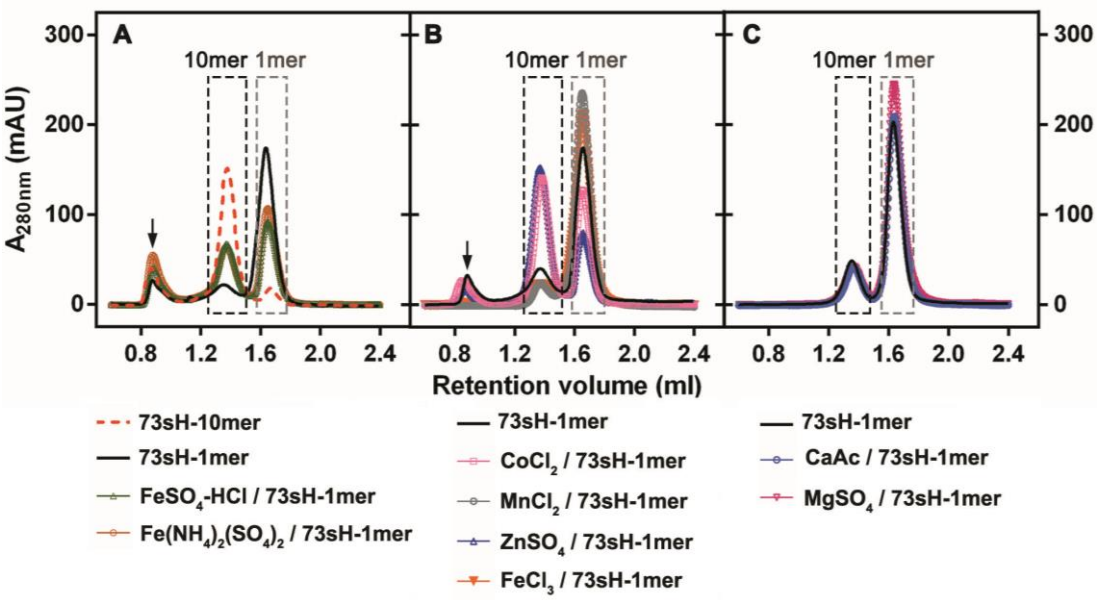


Figure 3.5.4 Analytical gel-filtration analysis on metal ions affecting the oligomeric states of Rru-EncFtnSH in solution.

90 µM Rru-EncFtnSH monomer fractions (73sH-1mer) were mixed with molar equivalent metal ion at room temperature for 2 hrs prior to analytical gel-filtration analysis. These metal solutions were freshly prepared including $FeSO_4$ in 0.1% (v/v) HCl, $Fe(NH_4)_2(SO_4)_2$, $CoCl_2$, $MnCl_2$, $ZnSO_4$, CaAc, $MgSO_4$ and $FeCl_3$. 90 µM Rru-EncFtnSH decamer fractions (73sH-10mer) were analysed via AGF as a control. Black arrows pointing to the void volume suggest the aggregate formed in protein samples. (A) suggests that Rru-EncFtnSH decamerises in the presence of Fe(II). (B) highlights the decamerisation effect of Co(II) and Zn(II), while Fe(III) and Mn(II) could not decamerise Rru-EncFtnSH as an equal molar addition. (C) Ca(II) and Mg(II) had no effects on Rru-EncFtnSH decamerisation. This AGF experiment has been repeated twice using different batches of proteins, which has generated the similar pattern. One of the AGF experiments is presented here without protein replicates.

3.5.3 Disassembly of Rru-EncFtnSH by collision-induced dissociation (CID) tandem mass spectrometry

The gas phase disassembly of the decameric Rru-EncFtnSH was performed using collision-induced dissociation (CID) tandem mass spectrometry (CID-TMS). Under the correct CID conditions, protein assemblies can dissociate with retention of subunit and ligand interactions, and thus provide structurally-informative evidence as to the topology of the original assembly; this has been termed 'atypical' dissociation (Hall et al. 2013). For Rru-EncFtnSH, this atypical dissociation pathway was clearly evident; CID of the Rru-EncFtnSH decamer resulted in the appearance of a dimeric Rru-EncFtnSH subcomplex containing 0, 1, or 2 iron ions (**Figure 3.5.5**). In light of the crystal structure, this observation can be rationalised as dissociation of the Rru-EncFtnSH decamer by disruption of the non-FOC interface with at least partial retention of the FOC interface and the FOC-Fe.

The CID-TMS experiment provided an added-value to the PDBePISA server analysis of dimer interface interaction, revealing that FOC dimer interface is more stable than the non-FOC dimer interface because of the iron binding. This observation also supports the crystallographic assignment of the overall topology of Rru-EncFtnSH structure as a pentameric assembly of dimers with two iron ions located at the FOC dimer interface. In addition, this analysis provides evidence that the overall architecture of the complex is consistent in the crystal, solution and gas phases.

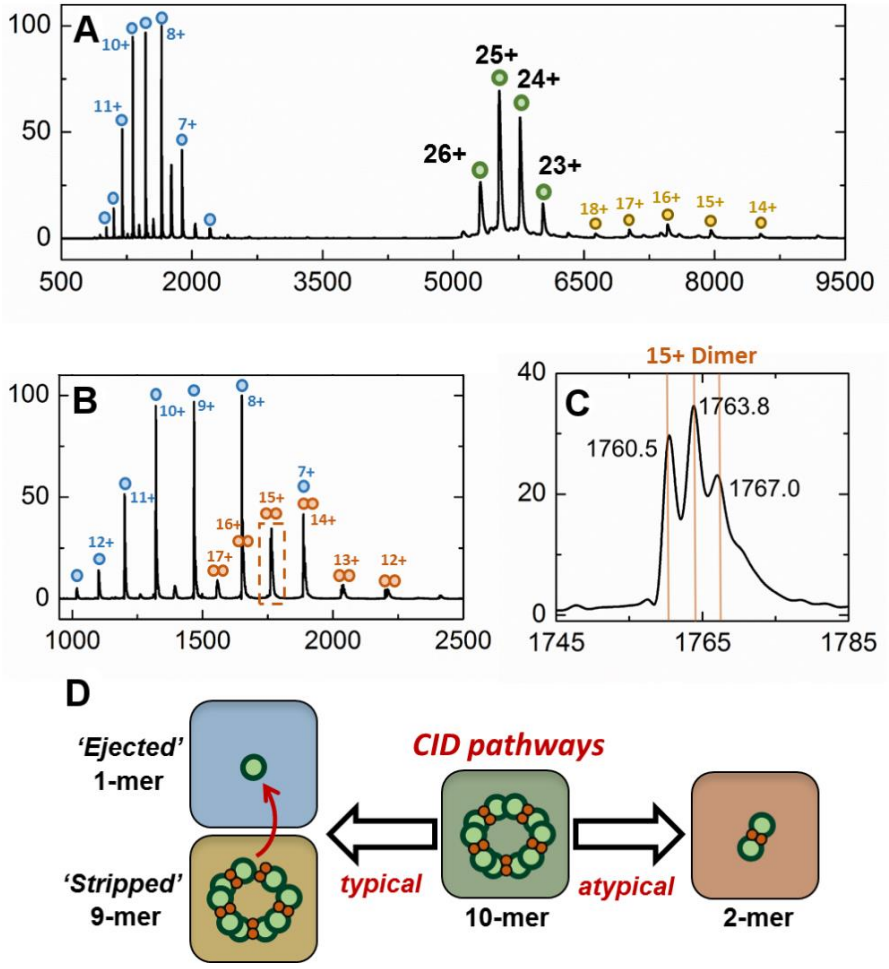


Figure 3.5.5 Gas-phase disassembly of the Rru-EncFtnSH decameric assembly.

The entire charge state distribution of the Fe-loaded decameric Rru-EncFtnSH assembly (green circles) was subject to collisional-induced dissociation (CID) by increasing the source cone voltage to 200 V and the trap voltage to 50 V. The resulting CID mass spectrum (A) revealed that dissociation of the holo-Rru-EncFtnSH decamer primarily occurred via ejection of a highly charged monomer (blue circles), leaving the 'stripped' complex (a 9mer; 118.7 kDa; yellow circles). The mass of the ejected-monomer is consistent with apo-Rru-EncFtnSH (13.2 kDa), suggesting unfolding of the monomer (and loss of Fe) occurs during ejection from the complex. This observation of asymmetric charge partitioning of the sub-complexes with respect to the mass of the complex is consistent with the "typical" pathway of dissociation of protein assemblies by CID (Hall et al. 2013). In addition, a third, lower abundance, charge state distribution is observed which overlaps the Rru-EncFtnSH ejected monomer charge state distribution; this region of the spectrum is highlighted in (B). This distribution is consistent with an ejected Rru-EncFtnSH dimer (orange circles). Interestingly, closer analysis of the individual charge state of this dimeric CID product shows that this sub-complex exists in three forms – displaying mass consistent with an Rru-EncFtnSH dimer binding 0, 1, and 2 Fe ions. This is highlighted in (C), where the 15+ charge state of the Rru-EncFtnSH dimer is shown; three peaks are observed with *m/z* 1760.5, 1763.8, and 1767.0 Th—the lowest peak corresponds to neutral masses of 26392.5 Da [predicted Rru-EncFtnSH dimer, $(C_{572}H_{884}N_{172}O_{185}S_2)_2$—26388.6 Da]. The two further peaks have a delta-mass of ~+50 Da, consistent with Fe binding. These are observed as partial 'atypical' CID fragmentation of the decameric complex—i.e. fragmentation of the initial complex with retention of subunit and ligand interactions. A schematic summary of these results is displayed in (D). The high stability of this iron-bound dimer sub-complex is largely dependent on the metal coordination at the dimer interface, increasing the strength of the dimer interface. Taken together, these observations support our findings that the topology of the decameric Rru-EncFtnSH assembly is arranged as a pentamer of dimers, with two Fe ions at each dimer interface.

3.6 Summary and discussion

Recombinant Rru-EncFtnSH existed as both monomer and decamer in solution; and these different forms of the protein differed in the amount of iron associated with the protein. Both monomer and decamer fractions crystallised into an annular decamer structure with an FOC dimer interface that coordinated di-iron in the FOC and calcium in metal ion entry/exit sites, as well as non-FOC dimer interface without metal binding. Iron ions were able to induce the decamerisation of Rru-EncFtnSH based on the iron supplementation to protein or medium. Analytical gel-filtration and native PAGE experiments confirmed that Fe(II) rather than Fe(III) induced the assembly of Rru-EncFtnSH, suggesting the high binding specificity of Rru-EncFtnSH towards Fe(II). Further metal titration experiments showed that Co(II) and Zn(II) were also able to induce the decamerisation of Rru-EncFtnSH possibly due to the similar coordination environment with iron ion. The Fe(II) titration into monomeric Rru-EncFtnSH instantly followed by nESI-MS experiment, only showed the decreasing monomer and increasing decamer. The absence of dimer from the MS spectra suggested that the formation of FOC dimer is rate-determining step. The IM-MS experiment indicated that the Rru-EncFtn monomer was partially-unstructured while the decamer was a compact structure with mainly α-helical character. CID-TMS disassembly experiments indicated that the FOC dimer interface was more stable than non-FOC interface.

Apart from the influence of metals on Rru-EncFtnSH assembly, Rru-EncFtnSH also displayed time-dependent equilibrium between monomer and decamer. The concentration-dependent shift of oligomeric state has only been observed between monomer and dimer. Given that no metal was supplemented, it was very likely that the high concentration favours the formation of non-FOC dimer; or the protein was soaking up trace metal from the solution in a very slow rate.

Taken together, it was postulated that under physiological conditions Rru-EncFtnSH assembled from monomer to FOC dimer due to the presence of Fe(II); once assembled, a pentamer of FOC dimers formed a ring-like decamer structure. It was also possible that the monomeric fractions, at high concentration during crystallisation, formed non-FOC dimer first. Then it scavenged trace amounts of iron into its FOC. Once the FOC

was formed, the decamer was assembled alongside the occupation of the metal ion entry/exit sites with calcium from the crystallisation conditions.

Chapter 4 Structures of *R. rubrum* EncFtn mutants

4.1 Introduction

The crystal structure of Rru-EncFtnSH revealed three metal binding sites: FOC (E32, E62, H65, Y39), metal ion entry site (E31, E34) and metal ion exit site (H57, E61 and E64). To understand how these metal binding sites influenced the structure, assembly, and biochemistry of Rru-EncFtnSH, several point mutants in the FOC and metal ion entry site were designed based on residues coordinating the bound metal ions. These mutants were constructed in the pET-28a-Rru-EncFtnSH and pACYCDuet-1-Rru-EncFtn-Rru-Enc expression vectors. The Rru-EncFtnSH mutants were heterologously expressed in *E. coli* BL21(DE3) for studies on protein structure, assembly and biochemistry. The co-expressed Rru-EncFtn and Rru-Enc mutants were also generated for biochemical studies; however, these will be investigated in future work.

The FOC mutants include single mutations: E32A, E62A, H65A; double mutation: E32A-E62A; and triple mutation: E32A-E62A-H65A (3A). Single mutations, E32D and E62D, were also constructed to determine the metal binding specificity of the FOC as these mutants would provide more space for metal association due to the shorter side chain of aspartic acid residue than the glutamic acid residue. The metal ion entry site mutants include E31A and E34A. The side chain of W38 was not seen in the electron density map of the Rru-EncFtnSH WT structure possibly due to its flexibility. Since W38 is involved in the FOC dimer interface and is spatially close to E34, it is very interesting to understand how the mutation of W38 affects the protein assembly. Therefore, a single mutation W38G, double mutation E31A-W38G and triple mutation W38G-E31A-E34A (WEE) were prepared for protein assembly study. In this chapter, the assembly of Rru-EncFtnSH with FOC mutations was also studied to reveal how these changes affected the protein stability.

4.2 Purification and crystallisation of Rru-EncFtnSH mutants

All pET-28a-Rru-EncFtnSH based mutants were purified from *E. coli* cells grown in LB medium by IMAC and SEC, followed by crystallisation using methods described in **Section 2.5**.

4.2.1 Purification of Rru-EncFtnSH based FOC mutants

In **Figure 4.2.1**, the SEC chromatograms of Rru-EncFtnSH bearing single FOC mutations, E32A and E62A, displayed a decamer peak, while the H65A mutant displayed a monomer peak, indicating that the H65A mutation disrupted the formation of the Rru-EncFtnSH decamer. The single mutant E62D displayed a major decamer peak while E32D showed a monomer peak. Both double mutation E32A-E62A and triple mutation E32A-E62A-H65A displayed a broader chromatographic peak, suggesting that both mutants destabilised Rru-EncFtnSH decamer assembly. SDS-PAGE analysis of the purified E32A, E62A and E62D proteins displayed a monomer and a dimer band (**Figure 4.2.2 A/D/E**), which was reminiscent of the wild-type purification from LB medium or MM + Fe(II). This results suggest that E32A, E62A and E62D mutants did not significantly disrupt the protein assembly. Other mutants including E32D, H65A, E32A-E62A and E32A-E62A-H65A displayed a monomer band on SDS-PAGE (**Figure 4.2.2 B/C/F/G**), which was consistent with the SEC purification chromatograms and thus suggested that these four mutants have disrupted the formation of the decameric form of Rru-EncFtnSH. It is worth pointing out that the intensity of the dimer sized bands of these mutants were much lower than the counterparts of WT (**Figure 3.2.6**); this observation can be explained by the weakened FOC dimer interface because of the FOC mutation.

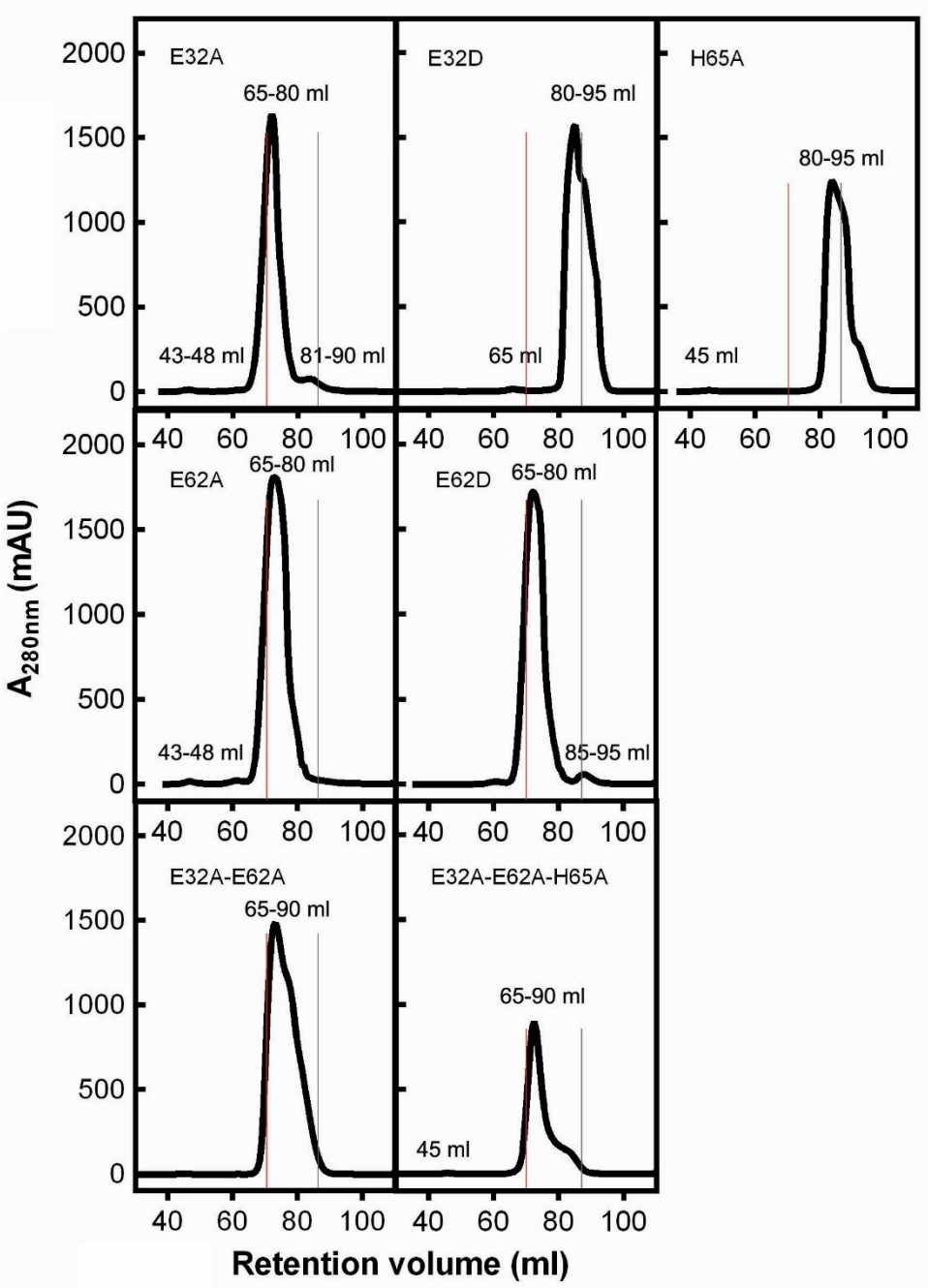


Figure 4.2.1 Size exclusion chromatograms of Rru-EncFtnSH FOC mutants purified from LB medium using HiLoad 16/600 Superdex 200 pg column.

FOC mutants including single mutations E32A, E32D, H65A, E62A, E62D, double mutation E32A-E65A and triple mutation E32A-E62A-H65A were purified through an S200 gel-filtration column. The red vertical lines highlight the peak position of WT decamer while the grey vertical lines highlight the position of WT monomer. Peak fractions were labelled as retention volume (ml) and analysed by SDS-PAGE in **Figure 4.2.2**. E32A, E62A and E62D display a major peak close to the WT decamer while E32D and H65A display a major peak close to the WT monomer. Double mutation E32A-E62A and triple mutation E32A-E62A-H65A display a broader chromatographic peak between WT decamer and monomer peak. The chromatographic observations suggest that E32A, E62A and E62D largely have kept the protein decamer assembly; E32D and H65A totally destablise the protein assembly, E32A-E62A and E32A-E62A-H65A make the protein partly disassociate.

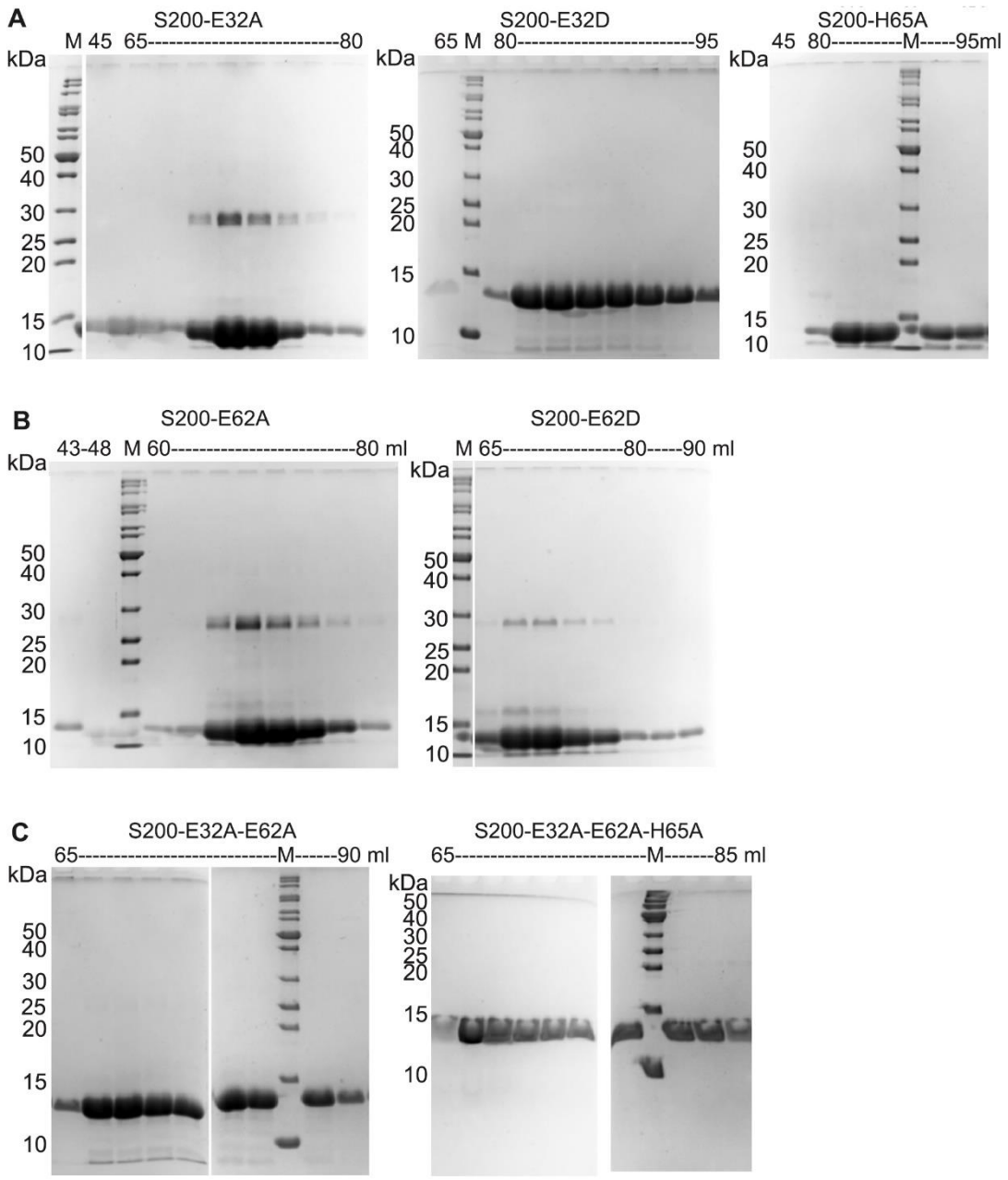


Figure 4.2.2 SDS-PAGE analysis of Rru-EncFtnSH FOC mutants purified via a SEC HiLoad 16/600 Superdex 200 pg column.

Peak fractions from SEC purified Rru-EncFtnSH mutants were subjected to 15% or 18% SDS-PAGE analysis: (A) E32A, (B) E32D, (C) H65A, (D) E62A, (E) E62D, (F) E32A-E62A and (G) E32A-E62A-H65A. E32A, E62A, E62D display two major bands on SDS-PAGE gels, corresponding to the monomer and dimer size of the mutants. E32D, H65A and double mutation E32A-E62A and triple mutation E32A-E62A-H65A only show monomer bands.

4.2.2 Mutants of Rru-EncFtnSH metal ion entry site

Four mutants of Rru-EncFtnSH were constructed in the metal ion entry site, including single mutations: E31A, E34A, W38G; and triple mutation: E31A-E34A-W38G. These variants of Rru-EncFtnSH were purified by IMAC and SEC, followed by SDS-PAGE analysis. The SEC chromatograms of four mutants all displayed a major decamer peak (**Figure 4.2.3**). The SDS-PAGE analysis of E31A and E34A mutants showed a mixture of monomer and dimer bands, in analogy to the wild-type protein. However, W38G and triple mutation E31A-E34A-W38G displayed only a monomer band. The peak fraction (B7) in the W38G SEC chromatogram showed a weak decamer-sized band aside from the monomer band. Therefore, it can be summarised that the single alanine mutation of E31 or E34 does not affect the protein assembly while the glycine mutation of W38—an FOC-dimer interfacing residue destablised the Rru-EncFtnSH assembly. The disassembly effect of W38G mutation has been enhanced by the triple mutation E31A-E34A-W38G.

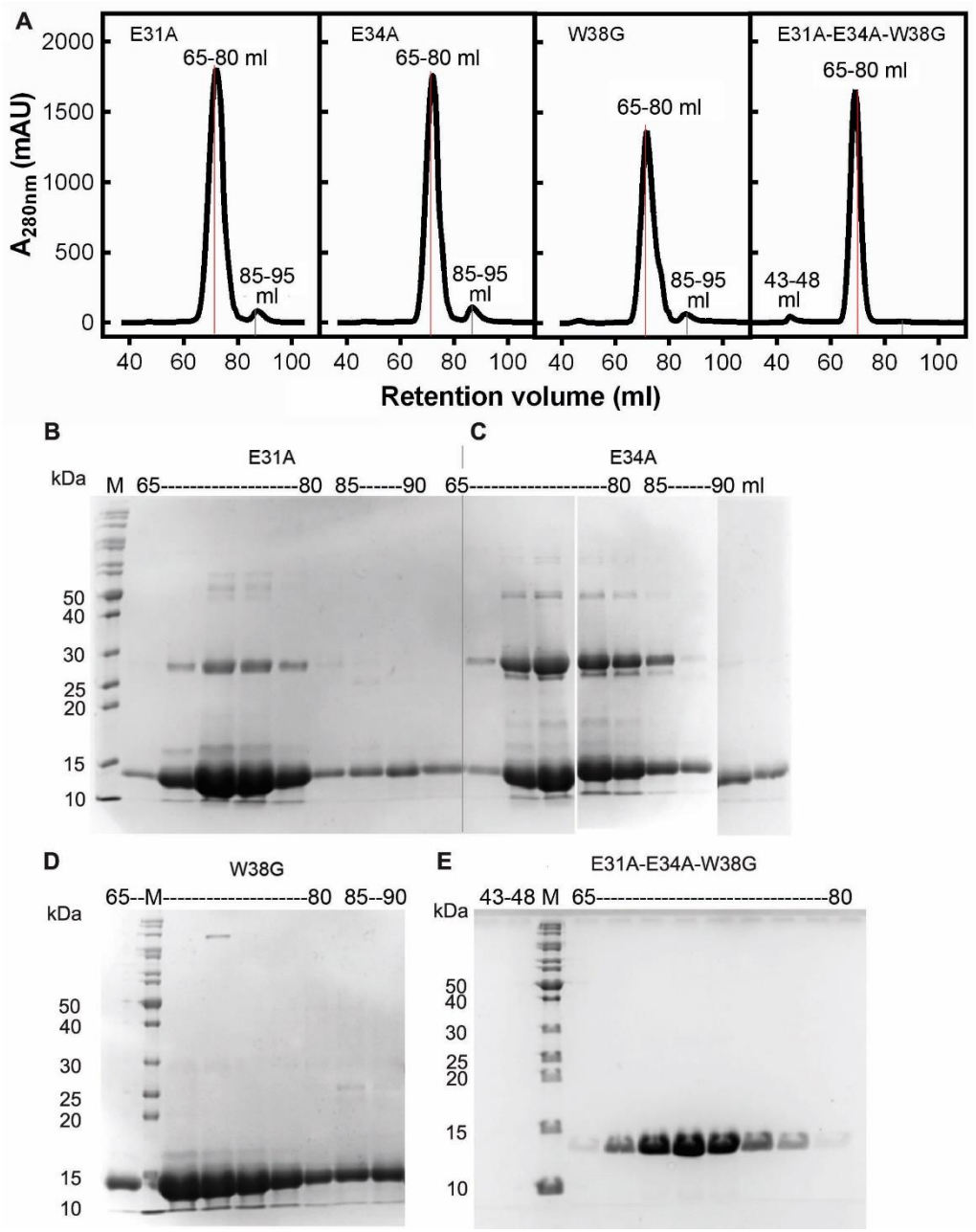


Figure 4.2.3 Size exclusion gel-filtration purification of Rru-EncFtnSH entry site mutants using HiLoad 16/600 Superdex 200 pg column.

(A) Chromatograms of Rru-EncFtnSH-E31A, E34A, W38G and triple mutation E31A-E34A-W38G subjected to SEC are shown. The red vertical lines highlight the peak position of WT decamer while the grey vertical lines highlight the position of WT monomer. All of these chromatograms display a major decamer peak and a minor monomer peak, suggesting that these mutants appear not to affect the protein stability. Fractions from the SEC peaks were subjected to 15% SDS-PAGE analysis. (B) E31A, (C) E34A, (D) W38G and (E) E31A-E34A-W38G. The SDS-PAGE analysis of E31A and E34A mutants show at least two major bands, corresponding to monomer and dimer. The W38G and triple mutation E31A-E34A-W38G both display a major single band around monomer size. These results suggest that the single mutations E31A and E34A does not alter the protein assembly, while the W38G mutation destablises the formation of protein oligomers.

4.2.3 Crystallisation results of Rru-EncFtnSH mutants

All of the Rru-EncFtnSH mutants purified from LB medium were concentrated to around 10 mg/ml and were crystallised under similar conditions to the wild-type protein, with the exception of double mutation E32A-E62A and triple mutation E32A-E62A-H65A, which crystallised from PEG 3350 with magnesium formate rather than calcium acetate (**Table 4.2.1**). Images of the crystals formed by these proteins are shown in **Figure 4.2.4**. X-ray data of all mutants have been collected without iron soaking except for E62D and E34A mutants. X-ray diffraction data from each of the crystals were collected at the 10–100 eV above the experimentally determined Fe-*K*α edge, except the 3A mutant (E32A-E62A-H65A), which was collected at 1.5 Å. The structures of each of the mutants were determined, except for W38G and W38G-E31A-E34A (WEE). The W38G crystal displayed poor quality of diffraction and thus was not refined. Although the WEE crystal diffracted up to 2.6 Å, the poor diffraction quality and crystal defects including twinning and multiple lattices have prevented the unambiguous assignment of the space group; hence this data were not refined further.

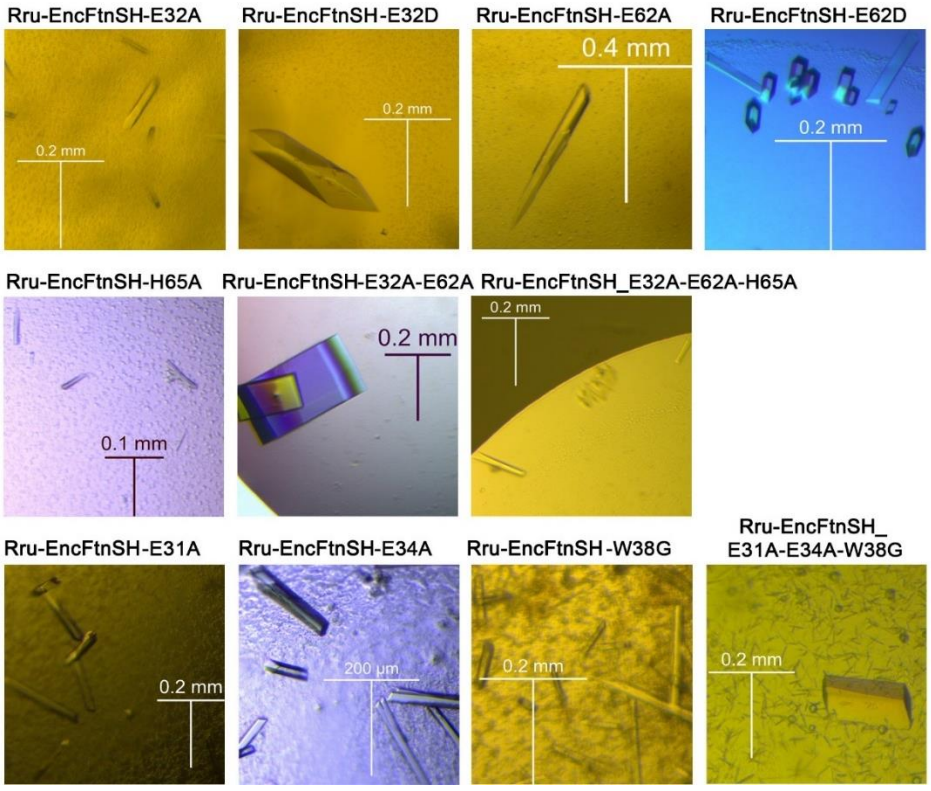


Figure 4.2.4 Crystal images of Rru-EncFtnSH mutants.

Horizontal and vertical scale bars are shown in each image to show the crystal size.

Based on pET-28a-Rru-EncFtnSH	Mutation site	Protein acronym	X-ray diffraction resolution	Space group	Crystallisation conditions	Cryo-protectant
Single mutation	FOC	E32A	2.74 Å	*P* 1 2_1 1	14% (w/v) PEG 3350 + 0.14 M CaAc	14% (w/v) PEG 3350 + 0.14 M CaAc + 20% (v/v) PEG 200
		E62A	2.34 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.13 M CaAc	15% (w/v) PEG 3350 + 0.13 M CaAc + 20% (v/v) PEG 200 + 1 mM acidic $FeSO_4$
		E32D	2.27 Å	*P* 1 2_1 1	14% (w/v) PEG 3350 + 0.14 M CaAc	14% (w/v) PEG 3350 + 0.14 M CaAc + 20% (v/v) PEG 200
		E62D	2.50 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.13 M CaAc	15% (w/v) PEG 3350 + 0.13 M CaAc + 20% (v/v) PEG 200+ 1mM acidic $FeSO_4$
		H65A	2.97 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.15 M CaAc	15% (w/v) PEG 3350 + 0.15 M CaAc + 20% (v/v) PEG 200
	Metal ion entry site	E31A	2.75 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.1 5M CaAc	15% (w/v) PEG 3350 + 0.1 5M CaAc + 20% (v/v) PEG 200
		E34A	2.19 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.14 M CaAc	15% (w/v) PEG 3350 + 0.14 M CaAc + 20% (v/v) PEG 200 + 50 mM Tris-HCl (pH 8.5) + 1 mM acidic $FeSO_4$
		W38G	3.32 Å	*P* 1 2_1 1	15% (w/v) PEG 3350 + 0.13 M CaAc	15% (w/v) PEG 3350 + 0.13 M CaAc + 20% (v/v) PEG 200
Double mutation	FOC	E32A-E62A	2.10 Å	*C* 1 2 1	0.2 M Magnesium formate + 18% (w/v) PEG 3350	0.2 M Magnesium formate + 18% (w/v) PEG 3350 + 17% (v/v) PEG 200
Triple mutation	FOC	E32A-E62A-H65A	2.20 Å	*P* 4_1 2_1 2	0.2 M Magnesium formate + 20% (w/v) PEG 3350	0.2 M Magnesium formate + 20% (w/v) PEG 3350 + 15% (v/v) PEG 200
	Metal ion entry site	W38G-E31A-E34A	2.64 Å	*P* 4_1 2_1 2	13% (w/v) PEG 3350 + 0.14 M CaAc	13% (w/v) PEG 3350 + 0.14 M CaAc + 20% (v/v) PEG 200

Table 4.2.1 Crystallisation results of Rru-EncFtnSH mutants.

The crystallisation related solutions are diluted from stock solutions, which includes 50% (w/v) PEG 3350, 100% (v/v) PEG 200, 1 M CaAc, 1 M magnesium formate, 1 M Tris-HCl (pH 8.5). 100 mM $FeSO_4$ is freshly dissolved in 0.1% (v/v) before diluting in the cryo-protectant solutions.

4.3 Assembly of FOC mutants

4.3.1 FOC mutants purified from MM ± Fe(II)

E32A, E62A, H65A and E32A-E62A-H65A (3A) were produced in *E. coli* cells grown in 1 litre minimal medium with or without Fe(II) and purified by IMAC and SEC. The size exclusion chromatograms are displayed in **Figure 4.3.1**. The E32A and E62A mutants eluted at a volume consistent with the decameric form of Rru-EncFtnSH, with a small proportion of monomer; the H65A mutant eluted at a volume consistent with the monomeric form of Rru-EncFtnSH. For the three single mutations of FOC studied, no change in oligomerisation states was observed upon addition of Fe(II) in the MM. These results suggest that the Fe(II)-dependent assembly of Rru-EncFtnSH was compromised by single mutations of FOC.

Interestingly, triple mutation E32A-E62A-H65A purified from 1 litre of MM ± Fe(II) showed an enhanced quantity of protein when compared to the other mutants, as well as chromatogram shifts toward higher oligomers compared to other three single mutations. The addition of Fe(II) in the MM of E32A-E62A-H65A led to a broader chromatographic peak, which may indicate some higher-order oligomer formation; or the enhanced production of protein in these growth conditions may have led to peak broadening as a consequence of overloading the column.

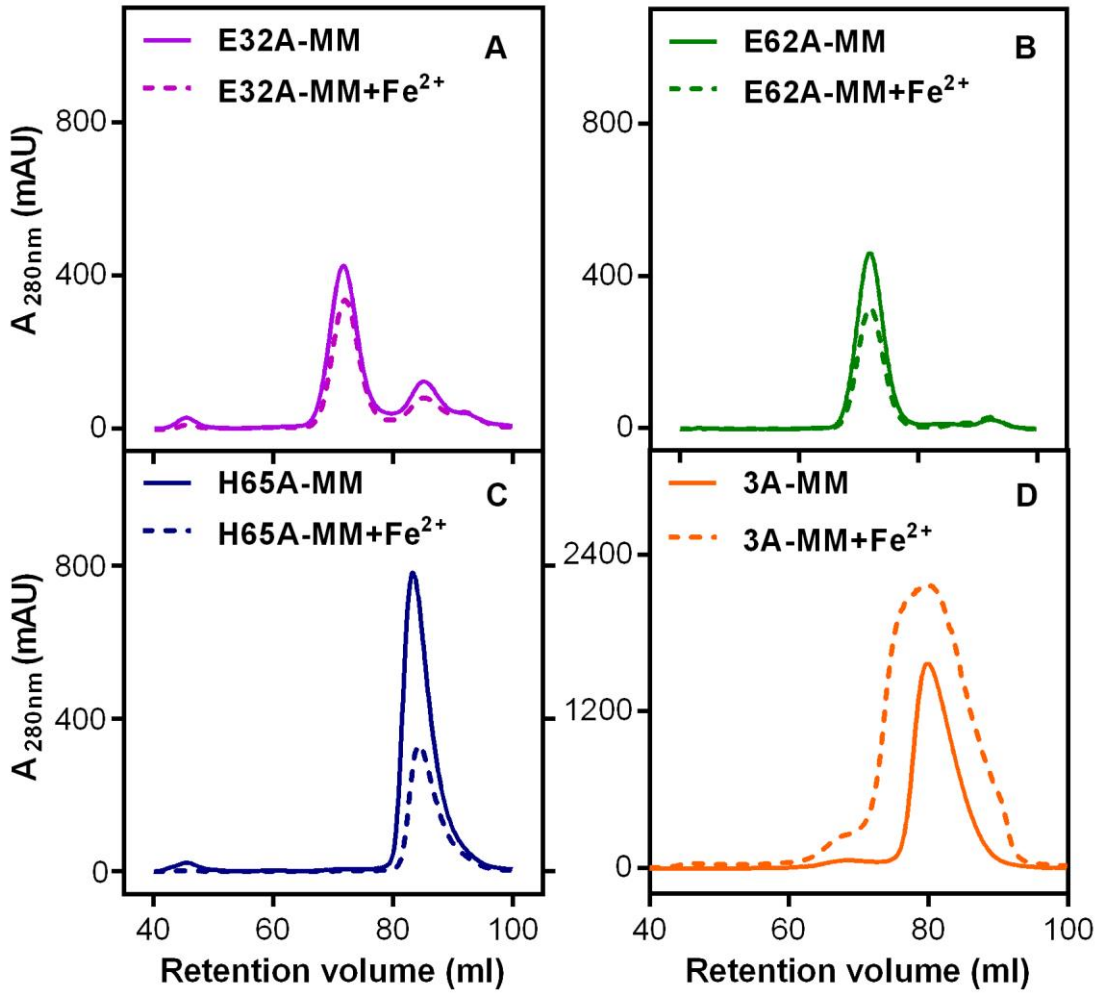


Figure 4.3.1 Size exclusion chromatograms of Rru-EncFtnSH FOC mutants purified from MM ± Fe(II) using HiLoad 16/600 Superdex 200 pg column.

Rru-EncFtnSH single mutations E32A, E62A, and H65A were produced from *E. coli* BL21(DE3) cells grown in MM or MM supplemented with Fe(II) were subjected to SEC. (A) Chromatograms of the E32A mutant purified from MM ± Fe(II) display a majority of the protein eluting as the decameric form of the protein and a small proportion of monomer. (B) Purified E62A mutant resulted in an elution profile with a major decameric peak. (C) Gel-filtration of the H65A mutant resulted in a single chromatographic peak corresponding to the monomer. (D) Gel-filtration of the triple mutation E32A-E62A-H65A (or 3A) purified from MM resulted in a major broad peak across several oligomeric states while the Fe(II) addition to the MM led to a broader chromatographic peaks.

4.3.2 The assembly of FOC mutants determined by native mass spectrometry

The major peak fractions of each mutant purified from MM were pooled and subjected to nESI-MS (or native mass spectrometry) analysis by Dr. David Clarke and Sam Huges (School of Chemistry, University of Edinburgh). The results were compared with the wild-type apo-Rru-EncFtnSH protein (**Figure 4.3.2**). According to our earlier results (**Figure 3.4.8**), the apo-Rru-EncFtnSH has a charge state distribution consistent with an unstructured monomer, and decamer formation is only initiated upon addition of ferrous iron. Both the E32A mutant and E62A mutant displayed charge state distributions consistent with decamers, even in the absence of Fe(II). This gas-phase observation is consistent with SEC measurements, which indicates both of these variants were also decamers in solution. Thus it appears that single mutation E32A or E62A allows the decamer assembly in the absence of iron in the FOC. In contrast to the glutamic acid mutants, MS analysis of the H65A mutant is similar to wild-type apo-Rru-EncFtnSH and is present as a monomer; interestingly a minor population of dimeric H65A was also observed. The triple mutation E32A-E62A-H65A, while showing primarily low-order assemblies in solution, displayed peaks consistent with decamer (10mer) and dodecamer (12mer) formation in the gas phase. The observation of the dodecamer has been reproducible, suggesting that instead of being an artefact the dodecamer of E32A-E62A-H65A could be a structural reassembly in gas phase or a random dimer attachment to the decamer ring structure.

The observed differences in the oligomeric state of the E32A and E62A mutants compared to wild-type possibly arise from the changes in the electrostatic environment within the FOC. At neutral pH the glutamic acid residues are negatively charged, while the histidine residues are predominantly in their uncharged state. In the wild-type Rru-EncFtnSH this leads to electrostatic repulsion between subunits in the absence of iron. Iron coordination in this site stabilises the dimer and reconstitutes the active FOC. The geometric arrangement of E32 and E62 in the FOC explains the behavior of the alanine substitution mutants in solution and the gas phase, where they both favor the formation of decamers due to the loss of a repulsive negative charge. The FOC in the H65A mutant is destabilised through the loss of this metal coordinating residue and potential positive charge carrier, thus favouring the monomer in solution and the gas phase. It

still remains elusive why the dodecamer appeared in gas phase study of E32A-E62A-H65A mutant.

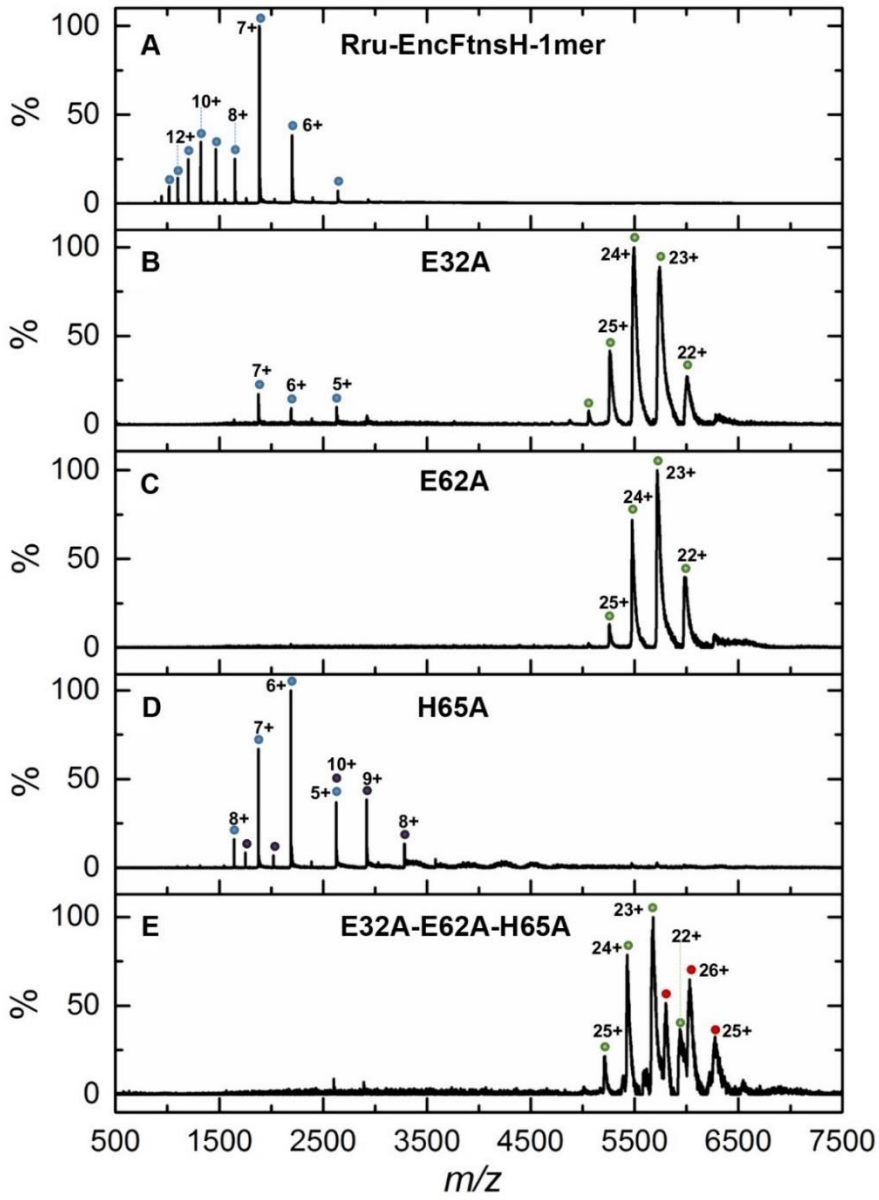


Figure 4.3.2 Native mass spectrometry of Rru-EncFtnSH WT and FOC mutants.

All spectra were acquired in 100 mM ammonium acetate, pH 8.0 with a protein concentration of 50 µM by Dr. David Clarke and Sam Huges. (A) WT (Rru-EncFtnSH monomer fractions) in the absence of iron displays a charge state distribution consistent with a monomer (see also figure 8). (B) E32A displays a charge states consistent with a decamer (green circles), a minor species, consistent with the monomer of E32A is also observed (blue circles). (C) E62A displays a charge states consistent with a decamer (green circles). (D) H65A displays charge states consistent with both monomer (blue circles) and dimer (purple circles). (E) apo-E32A-E62A-H65A displays two distinct charge states, consistent with a decamer (green circles) and a dodecamer (12mer; red circles).

4.4 Structures of Rru-EncFtnSH mutants

All Rru-EncFtnSH mutants purified from LB were subjected to crystallization trials. The X-ray crystal structures were determined in the similar way to the wild-type (WT) (**Section 2.5**), but using the monomeric or decameric WT structure as the search template in molecular replacement. The X-ray data collection and refinement statistics of all the mutants with structure solved are listed in **Table 4.4.1** and **Table 4.4.2**.

The crystal packing of all of the single mutation variants in the FOC of Rru-EncFtnSH in this study are essentially isomorphous to the WT structure, with three decamers in a *P* 1 2_1 1 space group. The E32A-E62A-H65A mutant displays a space group *P* 4_1 2_1 2 while the E32A-E62A mutant has a space group *C* 1 2 1; in both cases only a decamer is shown in the ASU. All of the mutants display the same decameric arrangement in the crystals as the wild-type Rru-EncFtnSH structure, and the monomers superimpose with an average $RMSD_{C\alpha}$ of less than 0.3 Å (**Figure 4.4.1**). A close inspection of residues involved in the metal binding site of each mutant highlights their effects on metal coordination.

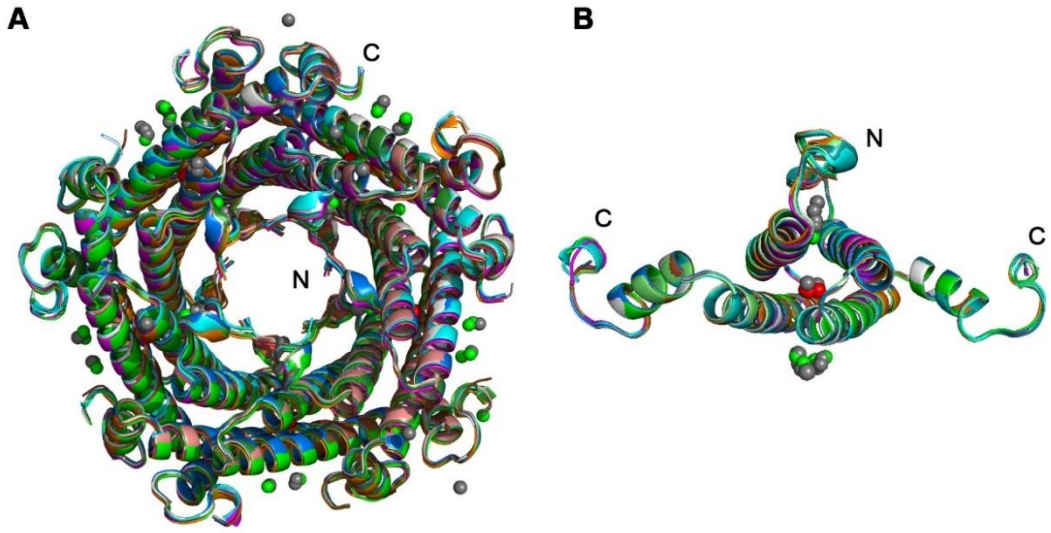


Figure 4.4.1 Structural superposition of FOC mutants and wild-type Rru-EncFtnSH.

(A) Aligned decamer structures of Rru-EncFtnSH wild-type (grey), mutants E32A (purple), E62A (green), H65A (cyan), 3A (orange), E32A-E62A-H65A (orange), E32A-E62A (brown), E62D (salmon), E31A (marine), E34A (forest) in cartoon mode with metal ion binding shown in spheres (iron-red, magnesium-green, calcium-grey). The positions of the N-termini and C-termini of the structures are marked. (B) Aligned FOC dimer structures of mutants and wild-type Rru-EncFtnSH are shown coloured as in (A).

	E32A	E62A	H65A	E32A-E62A-H65A
Data collection				
Beamline	I04	I02	I02	I04
Wavelength (Å)	1.73	1.73	1.74	1.5
Resolution range	48.84–2.59 (2.68–2.59)	48.87–2.21 (2.29–2.21)	48.86–2.97 (3.08–2.97)	48.41–2.25 (2.33–2.25)
Space group	*P* 1 2_1 1	*P* 1 2_1 1	*P* 1 2_1 1	*P* 4_1 2_1 2
Unit cell (Å) a	97.78	98.09	98.03	92.10
b	120.28	120.23	120.29	92.10
c	140.53	140.36	140.43	217.11
α = γ = 90°, ß =	95.41°	95.51°	95.39°	90°
Total reflections	405,488 (36,186)	1,069,345 (95,716)	323,853 (32,120)	26,1388 (23,040)
Unique reflections	100,067 (9,735)	162,379 (15,817)	66,658 (6,553)	44,951 (4,371)
Multiplicity	4.1 (3.7)	6.6 (6.1)	4.9 (4.9)	5.8 (5.3)
Completeness (%)	99 (97)	100 (97)	100 (99)	99 (99)
Mean I/sigma(I)	8.46 (1.79)	13.74 (1.80)	8.09 (1.74)	11.79 (1.72)
Wilson B-factor	40.1	33.97	52.01	34.6
R_{merge}	0.171 (0.792)	0.0979 (1.009)	0.177 (0.863)	0.112 (0.865)
R_{meas}	0.196 (0.923)	0.106 (1.107)	0.199 (0.966)	0.123 (0.961)
CC1/2	0.985 (0.557)	0.998 (0.642)	0.989 (0.627)	0.997 (0.557)
CC*	0.996 (0.846)	0.999 (0.884)	0.997 (0.878)	0.999 (0.846)
Image DOI	10.7488/ds/1419	10.7488/ds/1420	10.7488/ds/1421	10.7488/ds/1422
Refinement				
R_{work}	0.185 (0.292)	0.165 (0.299)	0.186 (0.271)	0.179 (0.264)
R_{free}	0.219 (0.350)	0.216 (0.364)	0.237 (0.325)	0.241 (0.318)
CC(work)	0.953 (0.796)	0.966 (0.751)	0.949 (0.816)	0.961 (0.782)
CC(free)	0.941 (0.647)	0.947 (0.670)	0.918 (0.814)	0.916 (0.661)
Number of non-hydrogen atoms	22362	22691	22145	7561
macromolecules	22005	21965	22066	7269
ligands	9	24	74	20
Protein residues	2684	2675	2700	898
RMS(bonds) (Å)	0.003	0.011	0.002	0.012
RMS(angles) (°)	0.53	1.02	0.4	1.08
Ramachandran favoured (%)	99	100	99	100
Ramachandran allowed (%)	0.8	0.42	0.91	0.34
Ramachandran outliers (%)	0	0	0.038	0
Rotamer outliers (%)	1.3	1.2	1.2	0.92
Clashscore	1.74	1.79	0.97	2.87
Average B-factor ($Å^2$)	41.25	41.34	47.68	41.05
macromolecules	41.29	41.31	47.6	41.07
ligands	73.18	65.55	72.34	46.01
solvent	37.74	41.46	33.85	40.09
Protein Data Bank code	5L89	5L8B	5L8G	5L8F

Table 4.4.1 X-ray data collection and refinement statistics of Rru-EncFtnSH mutants E32A, E62A, H65A and E32A-E62A-H65A.

Statistics for the highest-resolution shell are shown in parentheses. Friedel mates were averaged when calculating reflection numbers and statistics.

	E32A-E62A	E62D	E31A	E34A
Data collection				
Beamline	I04	I04	I24	I02
Wavelength (Å)	1.72	1.72	1.74	1.74
Resolution range	38.20–2.10 (2.17–2.10)	49.53–2.50 (2.59–2.50)	49.40–2.66 (2.75–2.66)	48.73–2.19 (2.27–2.19)
Space group	*C* 1 2 1	*P* 1 2_1 1	*P* 1 2_1 1	*P* 1 2_1 1
Unit cell (Å) a	140.82	97.97	97.89	97.79
b	66.43	120.27	120.16	120.14
c	117.46	139.85	139.37	139.34
α = γ = 90°, ß =	107.53	95.30	95.33	95.20
Total reflections	382,224 (36,130)	623,209 (56,386)	31,0118 (30,187)	90,6650 (86,276)
Unique reflections	58,549 (5,594)	111,676 (10,978)	9,0283 (8,770)	161,177 (15,444)
Multiplicity	6.5 (6.5)	5.6 (5.1)	3.4 (3.4)	5.6 (5.6)
Completeness (%)	0.96 (0.93)	1.00 (0.99)	0.97 (0.95)	0.97 (0.94)
Mean I/sigma(I)	13.73 (2.18)	9.60 (1.88)	7.55 (1.93)	11.27 (2.00)
Wilson B-factor	34.2	37.95	39.14	33.05
R_{merge}	0.0991 (0.794)	0.140 (0.737)	0.152 (0.700)	0.106 (0.761)
R_{meas}	0.108 (0.864)	0.155 (0.822)	0.181 (0.828)	0.117 (0.839)
CC1/2	0.998 (0.707)	0.994 (0.667)	0.985 (0.641)	0.997 (0.717)
CC*	0.999 (0.91)	0.999 (0.895)	0.996 (0.884)	0.999 (0.914)
Image DOI				
Refinement				
R_{work}	0.171 (0.243)	0.188 (0.283)	0.191 (0.296)	0.176 (0.277)
R_{free}	0.202 (0.273)	0.217 (0.304)	0.231 (0.316)	0.209 (0.313)
CC(work)	0.964 (0.868)	0.958 (0.808)	0.953 (0.806)	0.964 (0.839)
CC(free)	0.948 (0.786)	0.938 (0.786)	0.925 (0.768)	0.957 (0.760)
Number of non-hydrogen atoms	7698	22389	22386	22775
macromolecules	7239	22031	21988	22082
ligands	15	153	65	48
Protein residues	886	2668	2680	2698
RMS(bonds) (Å)	0.003	0.002	0.002	0.005
RMS(angles) (°)	0.48	0.46	0.44	0.62
Ramachandran favoured (%)	100	100	99	100
Ramachandran allowed (%)	0	0.46	0.8	0.45
Ramachandran outliers (%)	0	0	0	0.038
Rotamer outliers (%)	0.39	1.4	1.7	0.95
Clashscore	0.49	0.97	0.81	1.5
Average B-factor ($Å^2$)	37.95	44.21	41.07	39.69
macromolecules	37.7	44.21	41.13	39.69
ligands	47.27	47.77	53.51	45.77
solvent	41.76	42.05	34.94	39.35
Protein Data Bank code				

Table 4.4.2 X-ray data collection and refinement statistics of Rru-EncFtnSH mutants E32A-E62A, E62D, E31A and E34A.

Statistics for the highest-resolution shell are shown in parentheses. Friedel mates were averaged when calculating reflection numbers and statistics.

4.4.1 E32A structure

The FOC of the E32A mutant structure has no visible metal ions (**Figure 4.4.2**); the position of H65 is essentialy unchanged, while E62 moves closer towards where the di-iron centre would be in the wild-type structure (1.6–2.1 Å), so does Y39 (3.8–4.2 Å). But the absence of the axial-metal coordinating ligand provided by the E32 side chain abrogates metal binding in this site. The E32A mutant has five coordinated calcium ions among the fifteen metal ion entry sites (E31 and E34) in the ASU. The chi3 of some E31 residues are rotated by 180°, weakening the metal coordination in this site. Although all the side chains of H57, E61 and E64 resemble their counterparts in wild-type, there is no metal binding in the metal ion exit site of the E32A mutant.

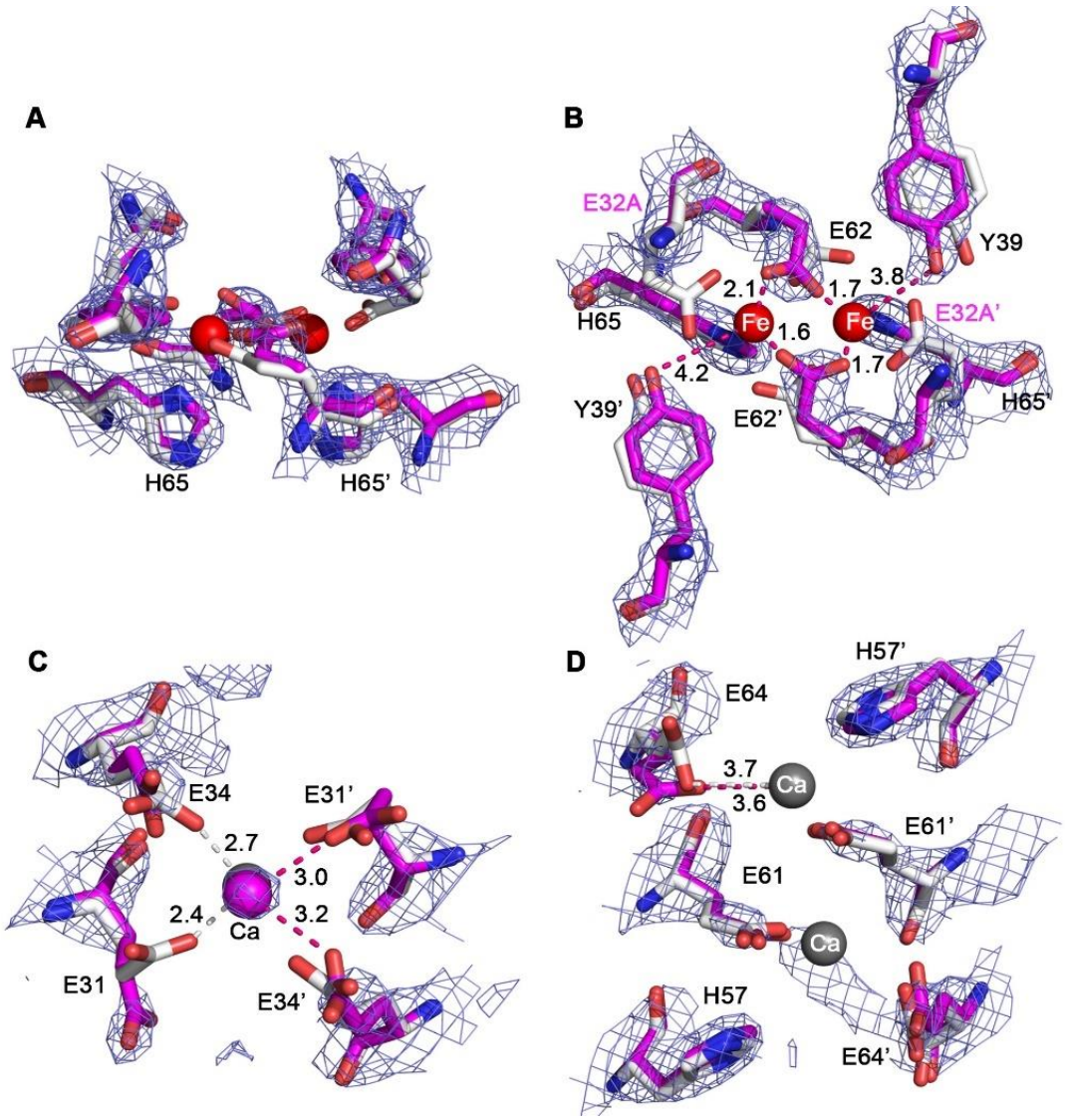


Figure 4.4.2 Comparison of metal binding sites in the E32A and WT Rru-EncFtnSH structures.

The metal binding sites in dimers of the E32A and WT Rru-EncFtnSH structures are shown in stick mode with Cα coloured in purple and white, respectively. The 2mFo-DFc density map for the mutant structure is superimposed on metal binding sites of E32A mutant as a blue mesh and contoured at 1.5 σ. Residues from dimers are differentiated with a prime symbol in their labels. (A) Aligned FOCs with di-iron (red spheres) coordinated in the WT structure. (B) Top view of (A) with residues labelled. The E32A mutant does not bind any metal in its FOC. (C) Metal ion entry site comparison between E32A mutant and WT. A single calcium ion is bound by both E32A mutant and WT in purple and grey spheres, respectively. (D) Metal ion exit site comparison between E32A mutant and WT with two calcium ions bound in grey spheres. The E32A mutant does not bind any metal in the exit site. The distance between iron or calcium ions and coordinating residues from E32A mutant, or WT, is marked by dashed lines and numbers nearby with Å as the unit.

4.4.2 E62A structure

In E62A mutant the position of the side chains of the remaining iron coordinating residues in the FOC is essentially unchanged (**Figure 4.4.3**). Since one carboxylate oxygen of E62 coordinates two iron in the FOC of WT, the alanine mutation of E62 abrogates metal binding in the FOC. The entry sites of E62A mutant in the ASU are all occupied with a calcium ion coordinated between residues E31 and E34. The metal ion exit site (H57, E61 and E64) coordinates zero to one calcium in the ASU. Among fifteen FOC dimers in the ASU of E62A structure, only two FOC dimers are occupied with a single calcium in the metal ion exit site. Both the metal ion entry site and exit site of the E62A mutant are analogous to the WT.

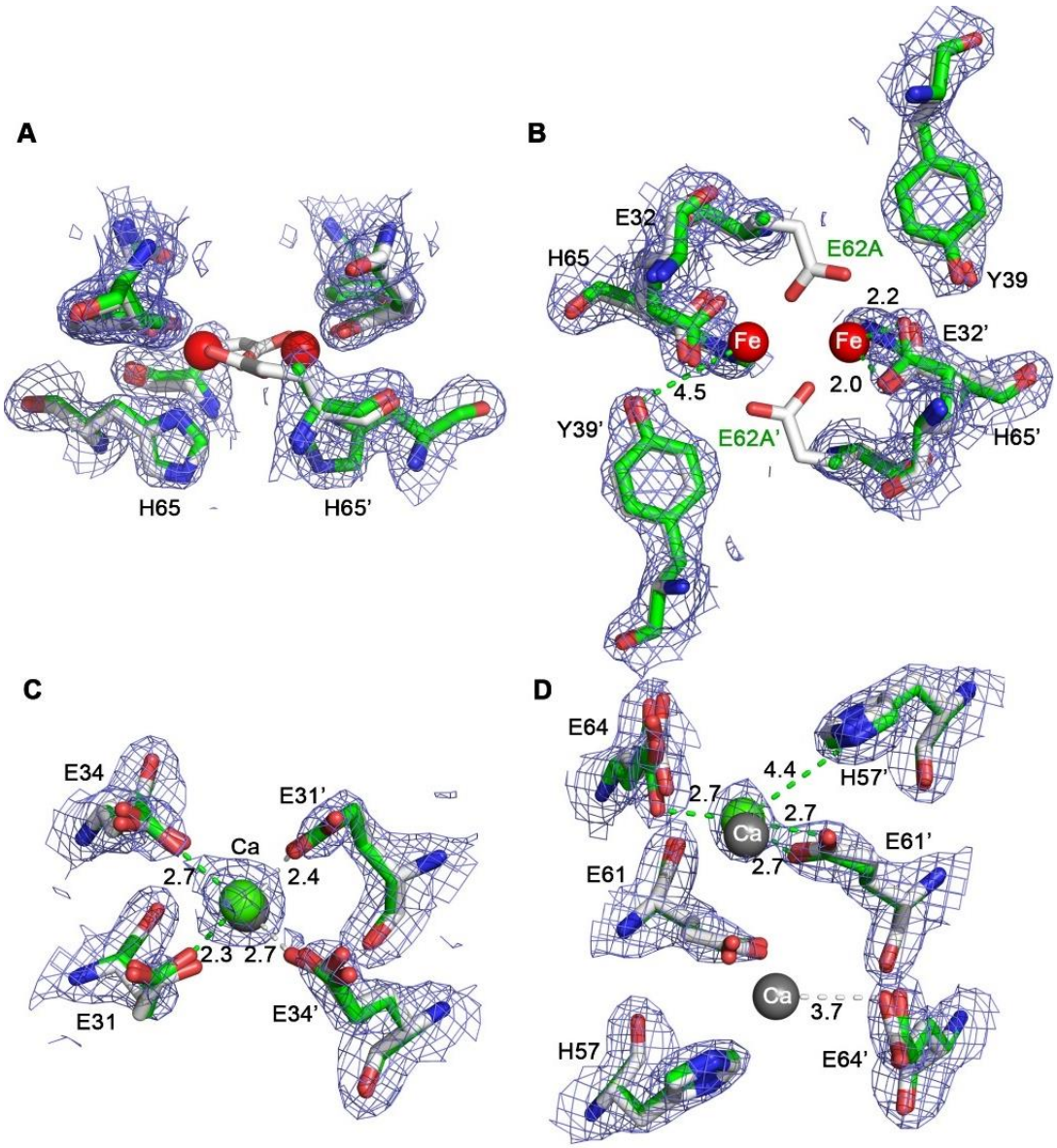


Figure 4.4.3 Comparison of metal binding sites in the E62A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E62A and WT are shown in stick mode with Cα coloured in green and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of E62A mutant as a blue mesh and contoured at 1.5 σ. Residues from dimers are differentiated with a prime symbol in their labels. (A) Aligned FOCs with di-iron (red spheres) coordinated in WT. (B) Top view of (A) with residues labelled. E62A mutant does not bind any metal in its FOC. (C) Metal ion entry site comparison between E62A mutant and WT. Ca ion is bound by both E62A mutant and WT in green and grey spheres, respectively. (D) Metal ion exit site comparison between E62A mutant and WT with two Ca ions bound in grey spheres. E62A mutant displays only one Ca ion (green sphere) at the exit site. The distance between iron or calcium ions and coordinating residues from E62A mutant or WT is marked by dashed lines and numbers nearby with Å as the unit.

4.4.3 H65A structure

H65A mutant coordinates two Ca(II) in the FOC, one Ca(II) in the entry site and two Ca(II) in the metal ion exit site (**Figure 4.4.4**). The H65A mutant diverges significantly from the wild-type in the position of the residues E32 and Y39 in the FOC. E32 appears in either the original orientation as the wild-type and coordinates Ca(II) in this position, or the chi3 is rotated by 104°, moving away from the coordinated calcium ion in the FOC. Y39 moves closer to Ca(II) compared to the wild-type and coordinates the calcium ion. Therefore, in FOC of H65A mutant, each calcium was coordinated by monodentate E62 and Y39, leading to a wider distance 5.1 Å between two calcium ions in the FOC. A single calcium ion was present in the entry site of this mutant; however, E31 of one chain is rotated away from the metal ion and is not involved in coordination. It is worth noting that because of the resolution limitation (3.0 Å) it is likely the conformation of E31 stayed the same as the WT. The metal ion exit site of H65A mutants is very similar to the WT.

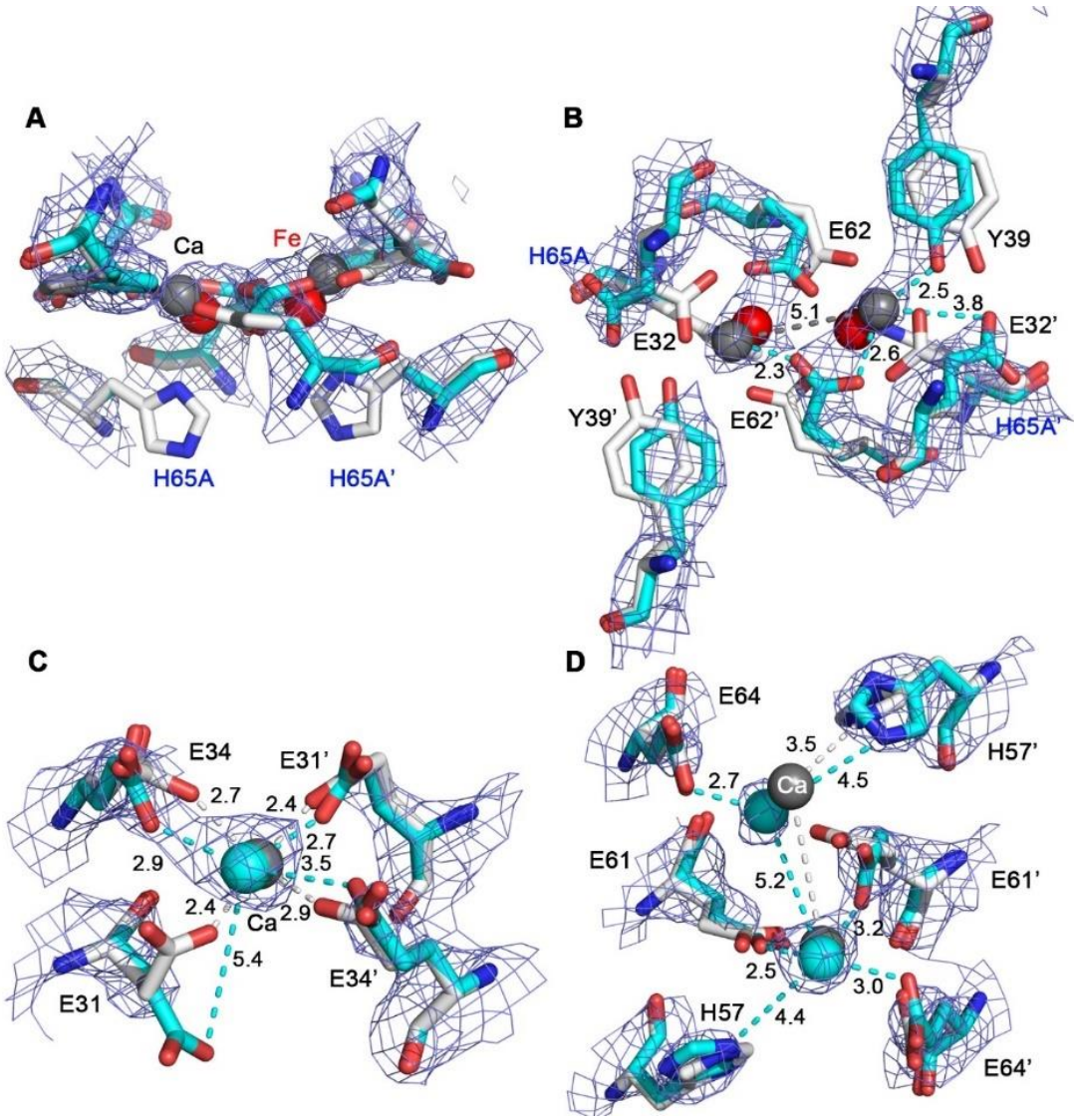


Figure 4.4.4 Comparison of metal binding sites in the H65A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric H65A and WT are shown in stick mode with Cα coloured in cyan and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of H65A mutant as a blue mesh and contoured at 1.5 σ. Residues from dimers are differentiated with a prime symbol in their labels. (A) Aligned FOCs with di-iron (red spheres) bound in WT and di-calcium (grey spheres) coordinated in H65A mutant. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between H65A mutant and WT. Ca ion is bound by both H65A mutant and WT in cyan and grey spheres, respectively. (D) Metal ion exit site comparison between H65A mutant and WT with two cyan and two grey spheres as Ca ions, respectively. The distance between iron or calcium ions and coordinating residues from H65A mutant or WT is marked by dashed lines and numbers nearby with Å as the unit.

4.4.4 E32A-E62A-H65A structure

When triple alanine mutations (E32A-E62A-H65A) are introduced in the FOC of WT, the FOC in the E32A-E62A-H65A mutant is totally abolished except for that only Y39 stays in the original conformation; and a triethylene glycol as the polyethylene glycol degradation product is coordinated in its original di-iron centre. No metal binds in the metal ion entry site of E32A-E62A-H65A mutant where E31 is not engaged in the metal coordination due to the rotation of E31 side chain. There are still two magnesium ions bound in the metal ion exit site, suggesting that the metal binding ability of metal ion exit site is not influenced by the mutation of FOC.

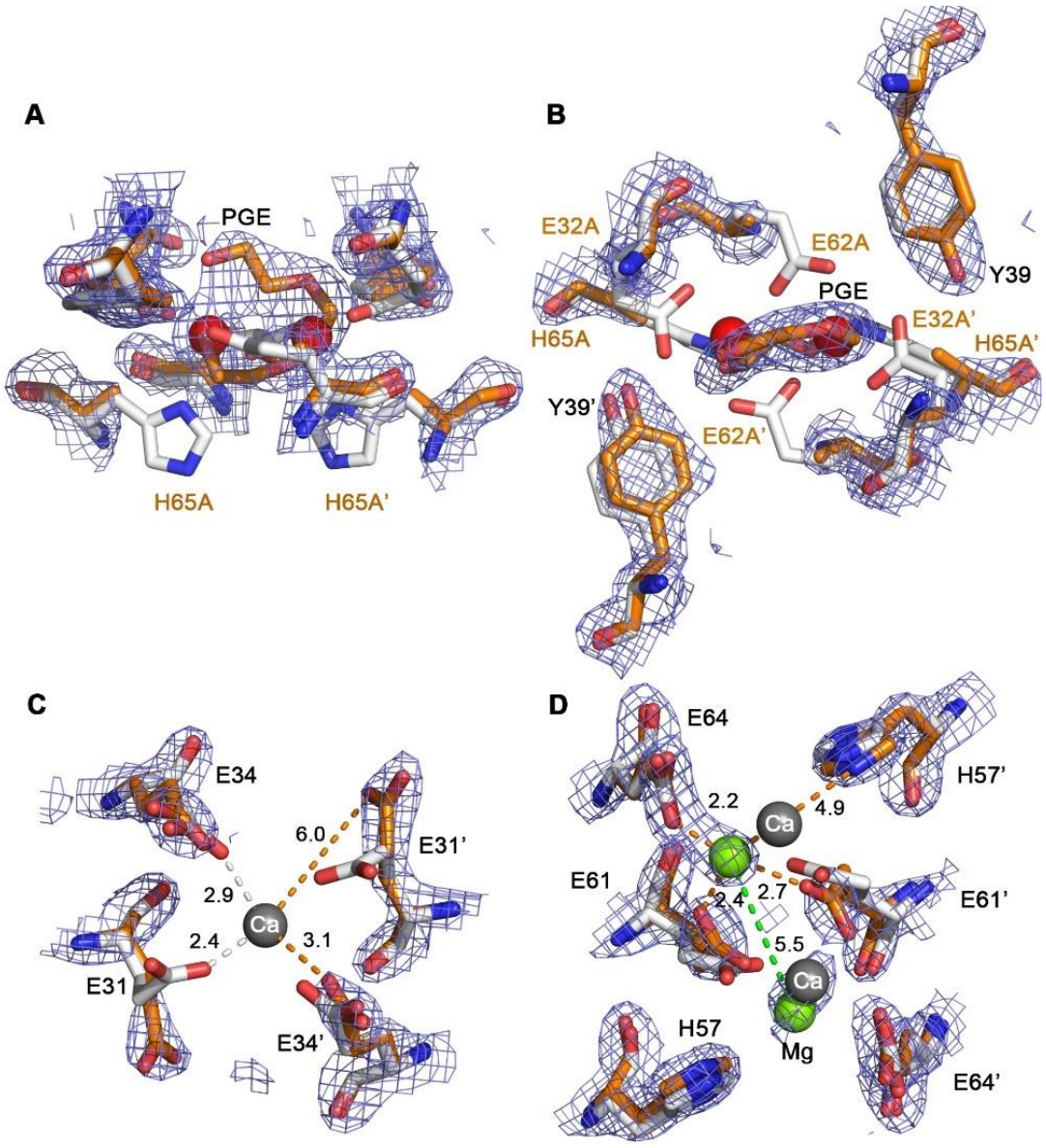


Figure 4.4.5 Comparison of metal binding sites in the E32A-E62A-H65A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E32A-E62A-H65A (3A) mutant and WT are shown in stick mode with Cα coloured in orange and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of 3A mutant as a blue mesh and contoured at 1.5 σ. Residues from dimers are differentiated with a prime symbol in their labels. (A) Aligned FOCs with di-iron (red spheres) bound in WT and triethylene glycol (PGE) coordinated in 3A mutant. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between 3A mutant and WT with Ca(II) in grey sphere. No metal is bound in the metal ion entry site of 3A mutant. (D) Metal ion exit site comparison between 3A mutant and WT. Two magnesium in green and two calcium ions in grey spheres are coordinated by 3A mutant and WT, respectively. The distance between iron, calcium or magnesium ions and coordinating residues from 3A mutant or WT is marked by dashed lines and numbers nearby with Å as the unit.

4.4.5 E32A-E62A structure

The E32A-E62A mutant structure displays a FOC without metal binding (**Figure 4.4.6**). Compared to the WT, the H65 conformation does not change; the Y39 moves from 4.7 Å to 4.0 Å towards the di-iron centre of the WT. No anomalous peaks were observed in this mutant structure although the diffraction datasets were collected at 10–100 eV above the theoretical Fe-*K*α edge, indicating the absence of iron binding in the E32A-E62A structure. Mg(II) instead of Ca(II) was modelled in the metal ion entry/exit sites of E32A-E62A mutant given that the crystallisation condition contained 0.2 M magnesium formate, which has been validated through the 'CheckMyMetal' web server (see the appendices) (Zheng et al. 2014). The Mg(II) in the metal ion entry site is coordinated by six water molecules and keeps a distance with E34 as 4.7 Å in contrast to the 2.9 Å in WT. The chi3 angle of E31 flips away from the Mg(II) by 116° and keeps a distance with Mg(II) at 5.5 Å. It suggested that the metal ion entry site of E32A-E62A mutant did not bind Mg(II); instead it was hexacoordinated Mg(II) bound by water present in the metal ion moving channel. There are two Mg(II) ions coordinated in the metal ion exit sites, which is far from with each other at 7.4 Å instead of 4.0 Å of Ca(II)-Ca(II) in the WT. Compared to the Ca(II) position, Mg(II) in this site moves closer to H57 from 4.5 Å to 2.1 Å and further from E64 of the adjacent chain from 3.2 Å to 4.1 Å. The Mg(II) in the metal ion exit site is also surrounded by four water molecules. Collectively, the double mutation of E32A and E62A abolished the metal binding in the FOC and affect the metal binding in the entry site, but the mutant was still able to coordinate metal ions in the exit site.

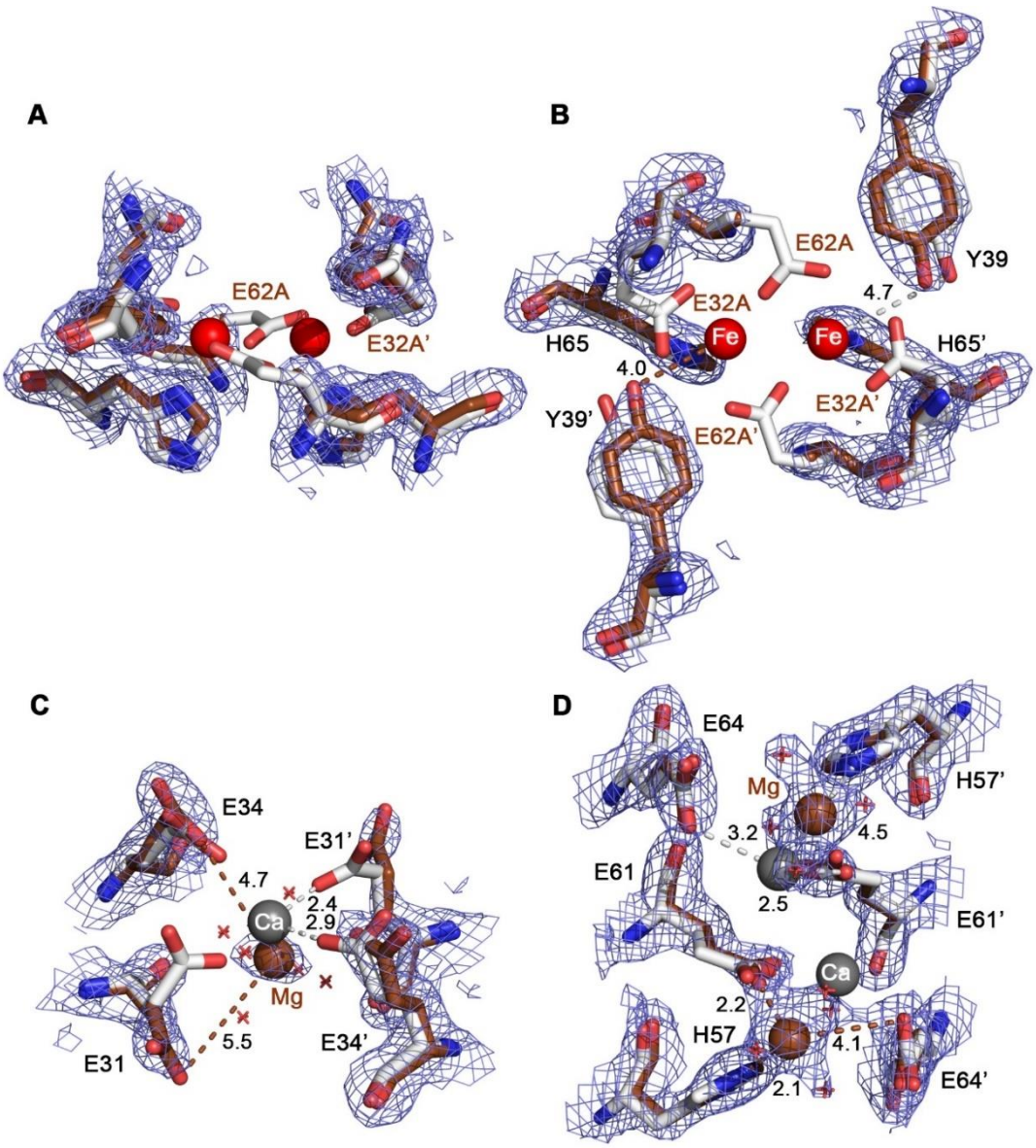


Figure 4.4.6 Comparison of metal binding sites in the E32A-E62A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E32A-E62A mutant and WT are shown in stick mode with Cα coloured in brown and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of E32A-E62A mutant as a blue mesh and contoured at 1.5 σ. Residues from dimers are differentiated with a prime symbol in their labels. The distance between metal ions and coordinating residues is marked by dashed lines and labelled with numbers nearby and Å as the unit. (A) Aligned FOCs with di-iron (red spheres) bound in WT. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between E32A-E62A mutant and WT with Mg and Ca ions bound in brown and grey spheres, respectively. The Mg(II) is coordinated by six water molecules shown in red crosses. The distances between Mg(II) and E31/E34 of the E32A-E62A mutant are 5.5 Å/4.7 Å by contrast to the 2.4–2.9 Å of the counterparts of WT. (D) Metal ion exit site comparison between E32A-E62A mutant and WT. Two Mg(II) ions shown in brown spheres in the mutant are bound by four water molecules in red crosses. Mg(II) is also coordinated by H57 in 2.1 Å and E61 in 2.2 Å. The distance between E64' and Mg(II) is 4.1 Å. Two Ca(II) ions shown in grey spheres in the WT are coordinated by H57, E61 and E64' in distances as 4.5 Å, 2.5 Å, 3.2 Å, respectively.

4.4.6 E62D structure

The E62D crystal has been soaked in 1 mM acidic $FeSO_4$ for 10–15 min prior to X-ray diffraction data collection at 10–100 eV above the theoretical Fe *K*-edge (**Figure 4.4.7**). The NCS-averaged anomalous peak height of 70 σ in the FOC confirmed the presence of di-iron ions. Apart from the E62D mutation, E32, H65 and Y39 kept the original conformation. The coordination of di-iron by the monodentate carboxylate oxygen was not affected by the E62D mutation. Above the di-iron ions a glycolic acid (GOA) was bound in the E62D in analogy to WT. In the metal ion entry site of E62D mutant statistically smaller NCS-averaged anomalous peaks were found at 40 σ and metal ion exit site at 25 σ, where Ca(II) ions were modelled. The metal ion entry site of E62D mutant consisted of a Ca(II) ion coordinated by E31 and E34 in slightly further distances than the WT, changing from 2.8 to 3.4 Å. The metal ion exit site of E62D mutant also consisted of two Ca(II) ions in a longer distance of 7.2 Å from each other. In contrast to the WT, the Ca(II) was bound by H57 in a shorter distance of 2.9 Å and a longer distance of 4.3 Å to E64'. Taken together, the E62D mutation altered the Rru-EncFtnSH structure and assembly to a minimal degree.

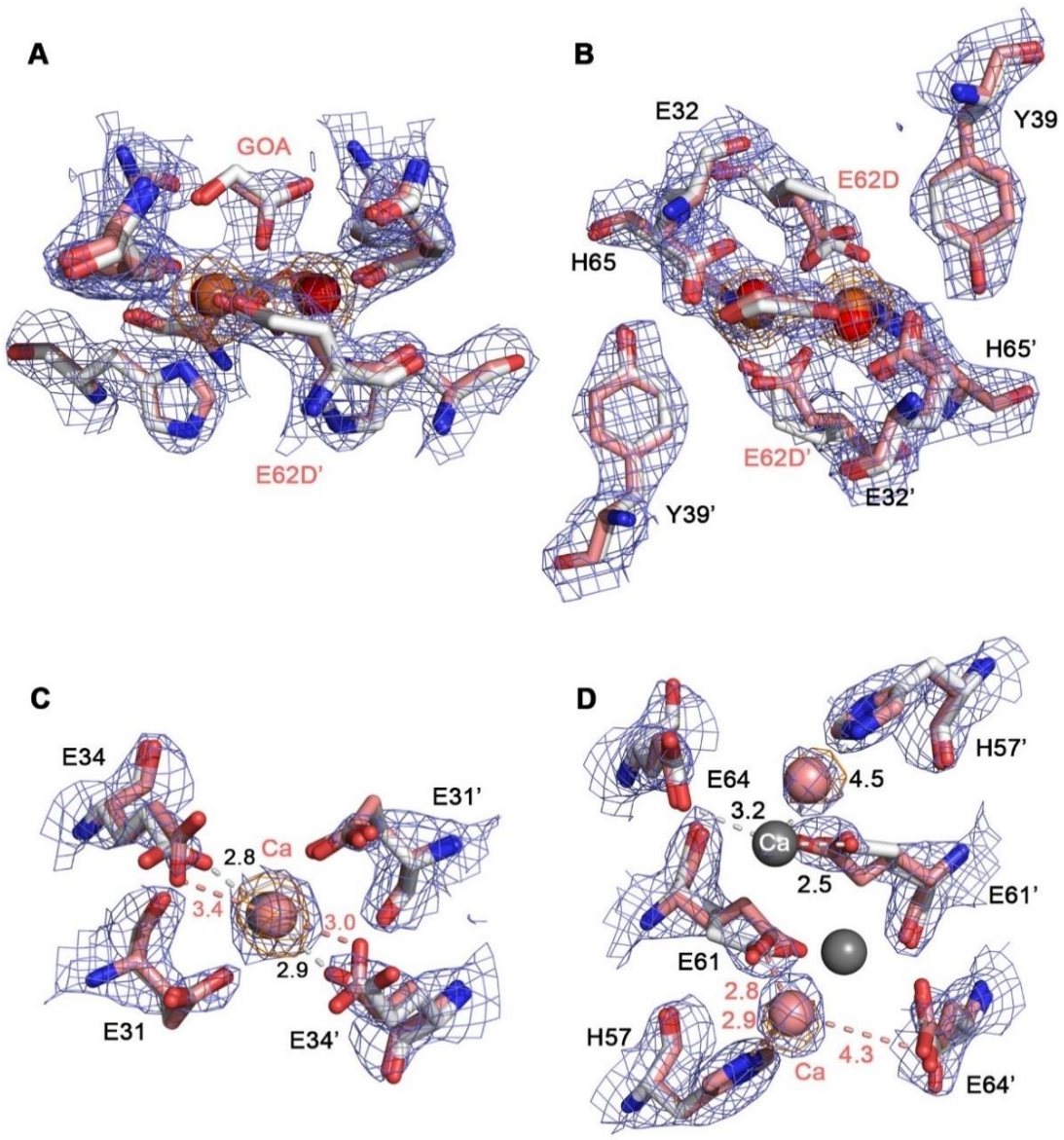


Figure 4.4.7 Comparison of metal binding sites in the E62D and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E62D and WT are shown in stick mode with Cα coloured in salmon and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of E62A mutant as a blue mesh and contoured at 1.5 σ. An NCS-averaged anomalous map is superimposed on FOC of E62D mutant as an orange mesh and contoured at 20 σ. Residues from dimers are differentiated with a prime symbol in their labels. (A) Aligned FOCs with di-iron coordinated as red spheres in WT and orange spheres in E62D mutant. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between E62D mutant and WT. Ca ion is bound by both E62D mutant and WT in salmon and grey spheres, respectively. (D) Metal ion exit site comparison between E62D mutant and WT with two Ca(II) ions bound in salmon and grey spheres, respectively. The distance between two Ca(II) ions from E62A mutant is 7.2 Å by contrast to 4.0 Å in the WT. The Ca(II) in E62D mutants keeps a distance with E61 as 2.8 Å, with H57 as 2.9 Å, with E64' as 4.3 Å.

4.4.7 E31A structure

An NCS-averaged anomalous peak height of 43 σ has been detected in the FOC of E31A mutant at the 10–100 eV above the theoretical Fe-*K*α edge, indicating the di-iron binding of FOC in E31A mutant. The E31A structure displays a similar FOC with WT with the exception of the water molecule bridging the di-iron instead of the glycolic acid in the WT structure (**Figure 4.4.8**). The metal ion entry site of E31A mutant still coordinated a Ca(II) by E34 from two chains in a distance of 3.4 Å/3.8 Å. The Ca(II) in the entry site of E31A structure stays further away from the di-iron binding in the FOC. The metal ion exit site of E31A mutant bound two Ca(II) in a distance of 5.5 Å with each other. The conformations of Ca(II) residues E61, H57 and E64 from the next chain kept the minimal change in the E31A mutant. The second Ca(II) in the metal ion exit site of E31A mutant moves closer to H57 from 4.9 Å to 3.1 Å and further to E64 of the adjacent chain from 3.1 Å to 3.8 Å. Overall, the E31A mutation did not alter the metal binding of Rru-EncFtnSH.

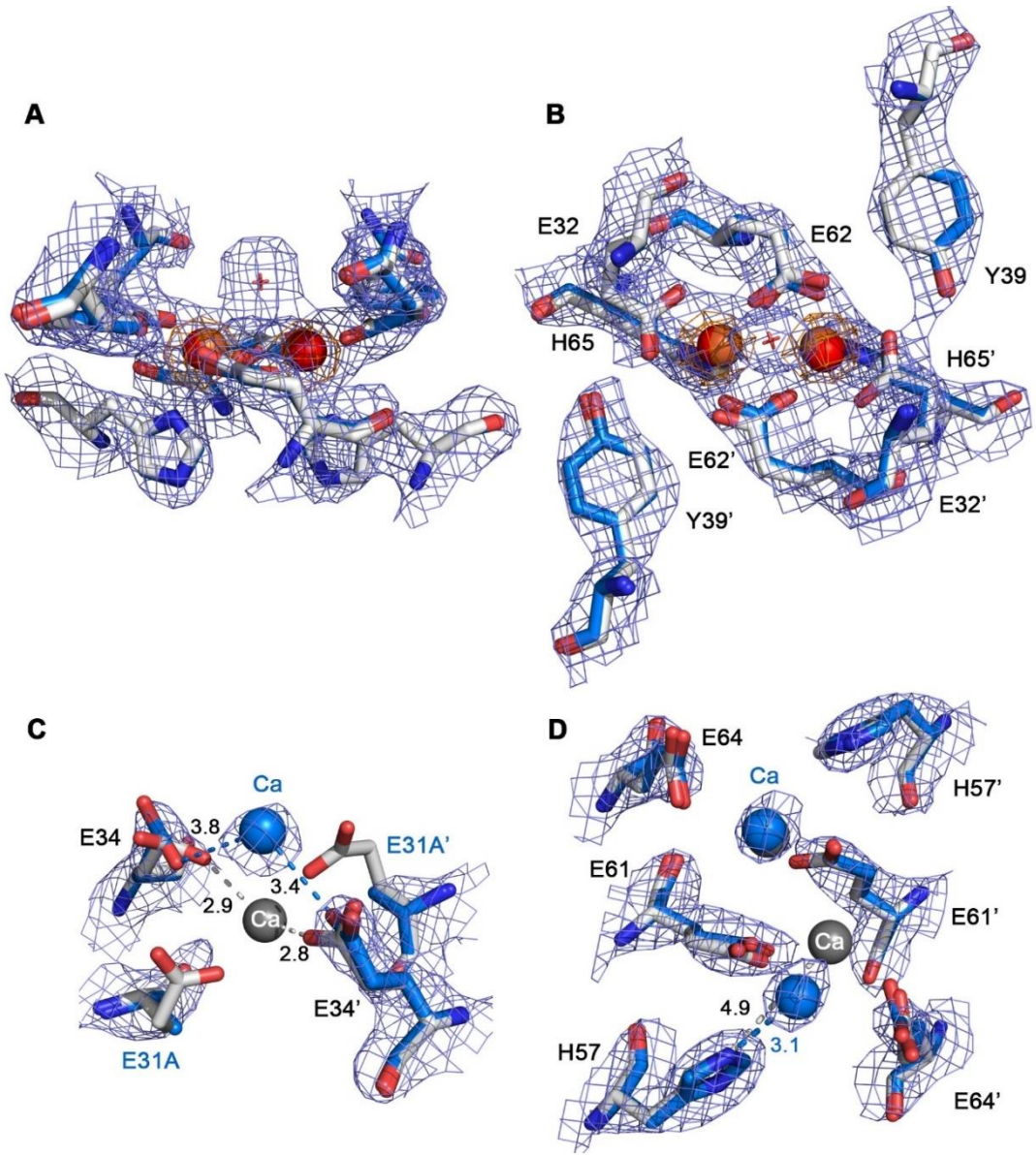


Figure 4.4.8 Comparison of metal binding sites in the E31A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E31A mutant and WT are shown in stick mode with Cα coloured in marine and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of E31A mutant as a blue mesh and contoured at 1.5 σ. An NCS-averaged anomalous map is superimposed on FOC of E31A mutant as an orange mesh and contoured at 20 σ. Residues from dimers are differentiated with a prime symbol in their labels. The distance between metal ions and coordinating residues is marked by dashed lines and labelled with numbers nearby and Å as the unit. (A) Aligned FOCs with di-iron bound as orange spheres in E31A mutant and red spheres in WT. A water molecule shown as a red cross is coordinated above the di-iron of E31A mutant. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between E31A mutant and WT with both Ca(II) bound in grey and brown spheres, respectively. In the E31A mutant the Ca(II) moves further from FOC and is coordinated by E34 from two chains in a distance of 3.4 Å/3.8 Å by contrast to the 2.8 Å/2.9 Å in the WT. (D) Metal ion exit site comparison between E31A mutant and WT, where two Ca(II) bound by both are shown in marine and grey spheres, respectively. One Ca(II) in E31A mutant moves closer towards H57 in 3.1 Å rather than 4.9 Å in the WT.

4.4.8 E34A structure

The crystal of E34A mutant was soaked in 1 mM $FeSO_4$ for 10–15 min prior to X-ray diffraction data collection at 10–100 eV above the theoretical Fe-*K*α edge. The structure of E34A mutant displayed a FOC with two NCS-averaged anomalous peaks at the height of 100 σ, indicating the di-iron ions binding in the FOC. All di-iron binding residues kept the minimal change compared the WT. No metal was observed in the metal ion entry site of E34A mutant, where the chi3 angle of E31 residue flipped away from the metal position because of the E34A mutation. The chi2 angle of W38 from the next chain flipped by 38° towards the metal position. The metal ion exit site of E34A mutant, in analogous to the WT, possessed a single Ca(II) whereas the second Ca(II) binding site was occupied by a water molecule. Given that in the WT structure the metal ion exit site was also partially occupied, the comparison between WT and E34A regarding this site displayed the minimal difference. Overall, the E34A mutation diminished the metal binding in the metal ion entry site but did not affect the FOC and metal ion exit site.

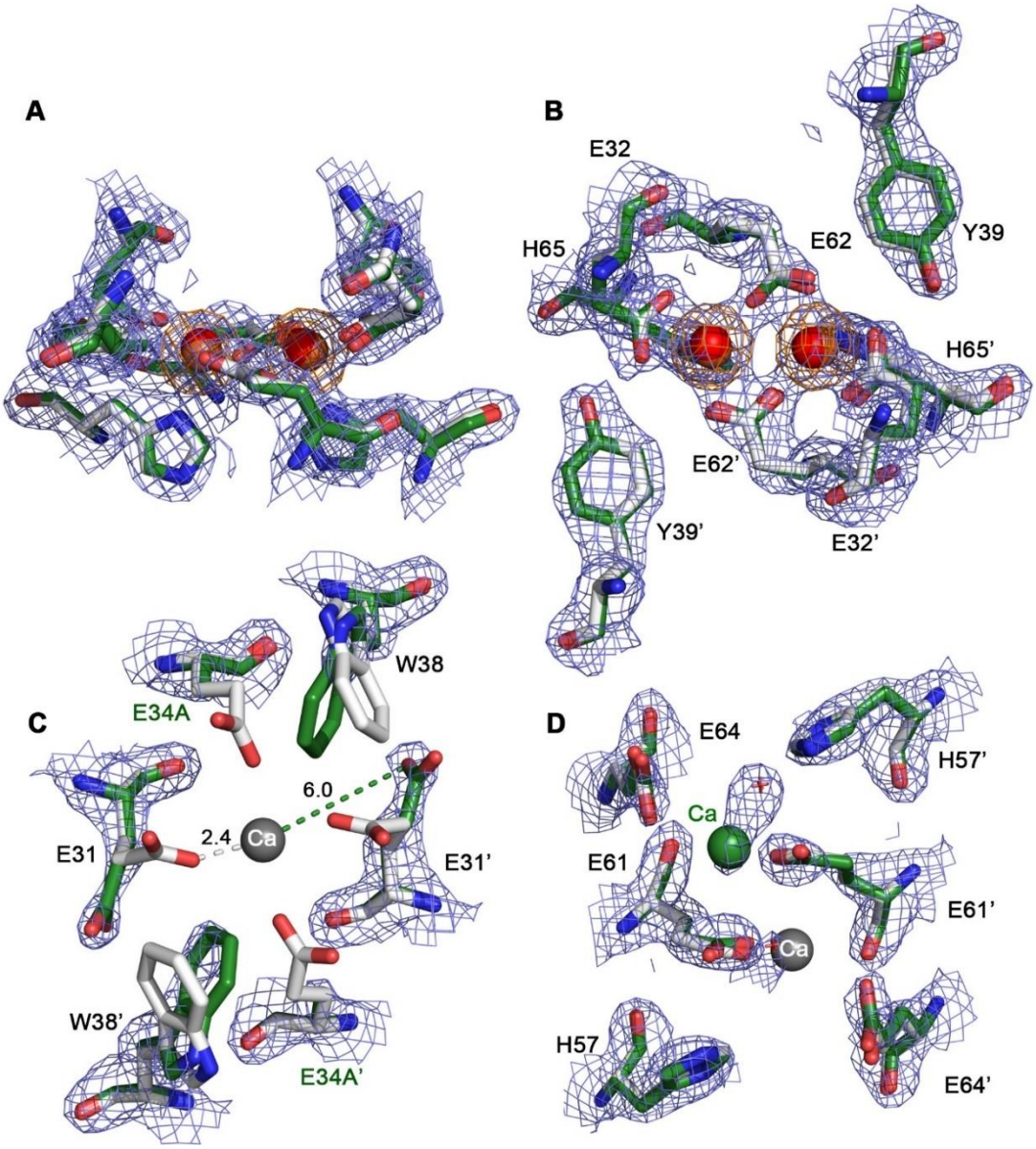


Figure 4.4.9 Comparison of metal binding sites in the E34A and WT Rru-EncFtnSH structures.

Metal binding sites in dimeric E34A mutant and WT are shown in stick mode with Cα coloured in forest and white, respectively. 2mFo-DFc density map is superimposed on metal binding sites of E34A mutant as a blue mesh and contoured at 1.5 σ. An NCS-averaged anomalous map is superimposed on FOC of E34A mutant as an orange mesh and contoured at 20 σ. Residues from dimers are differentiated with a prime symbol in their labels. The distance between metal ions and coordinating residues is marked by dashed lines and labelled with numbers nearby and Å as the unit. (A) Aligned FOCs with di-iron ions bound as orange spheres in E34A mutant and red spheres in WT. (B) Top view of (A) with residues labelled. (C) Metal ion entry site comparison between E34A mutant and WT. No metal is coordinated in this site of E34A mutant while WT binds a Ca(II) shown in grey sphere. The E31 in E34A mutant moves away by 117° and keeps a distance of 6 Å to the Ca(II) of WT. The indole ring of W38 rotates towards the central Ca(II) of WT by 38°. (D) Metal ion exit site comparison between E34A mutant and WT, where only one Ca(II) shown in forest sphere is bound by E34 while two Ca(II) in grey spheres are bound by the WT. The conformation of metal ion exit site does not change in the E34A mutant.

4.5 Summary and discussion

Based on Rru-EncFtnSH	Mutation site	Protein acronym	Oligomeric states in solution	Metal ion entry site	Near entry site	FOC	Metal ion exit site
Wild-type		WT	10mer+1mer	Ca(II)		2 Fe ions	2 Ca(II)
Single mutation	FOC	E32A	Mainly 10mer	Ca(II)		None	None
		E62A	10mer	Ca(II)		None	1 Ca(II)
		E62D	Mainly 10mer	Ca(II)		2 Fe ions	2 Ca(II)
		H65A	1mer	Ca(II)		2 Ca(II)	2 Ca(II)
	Metal ion entry site	E31A	Mainly 10mer	None	Ca(II)	2 Fe ions	2 Ca(II)
		E34A	Mainly 10mer	None		2 Fe ions	1 Ca(II)
Double mutation	FOC	E32A-E62A	Multiple	None	Mg(II)	None	2 Mg(II)
Triple mutation	FOC	E32A-E62A-H65A	10mer +12mer	None		PGE	2 Mg(II)

Table 4.5.1 Summary of protein assembly and metal binding in Rru-EncFtnSH mutants.

The oligomeric states of WT were determined using SEC-MALS. The oligomeric states of mutants E32A, E62A, H65A and E32A-E62A-H65A were estimated based on nESI-MS analysis. The oligomeric states of E62D and E32A-E62A were determined based on the calibration curve of the HiLoad 16/600 Superdex 200 pg column. In mutants E31A and E32A-E62A metal ion observed near the metal ion entry site, but were not bound by E31 and E34 residues in Ca/Mg coordination distance. In E32A-E62A-H65A mutant the FOC was occupied by triethylene glycol (PGE) instead of metal ions.

Each of the Rru-EncFtnSH mutants listed in **Table 4.5.1** crystallised into annular decamer structures in an analogous manner to the WT. The presence of additional Fe(II) in minimal medium did not alter the assembly of E32A, E62A, H65A and E32A-E62A-H65A variants, indicating that the alanine mutations of the FOC abrogated their iron binding abilities. This was later confirmed by the crystal structure of E32A, E62A and H65A, where the FOC was not populated by iron. The E62D displaying a similar metal binding with the WT, indicating that the E62D mutation did not alter the metal binding environment in the FOC of Rru-EncFtnSH. The E32A-E62A and E32A-E62A-H65A mutations both affected the metal binding in the metal ion entry site as well as the FOC. This is because the E32A mutation led to the conformational change

of E31 and thus abolished the metal coordination in the metal ion entry site, which was supported by the absence of metal in the entry site of E31A mutant. The E34A mutant also did not bind any metal in the entry site. Both E31A and E34A mutations did not affect the metal binding in the FOC and metal ion exit site.

To understand the function of Rru-EncFtnSH and the influence of the different metal coordination sites on the structure and activity of the protein, further biochemical analyses will be presented in **Section 5.5**. Since the E32D structure has not been finally determined, the structure together with the application in alternative metal coordination will be explored in future work. Besides, it is also very important to investigate the metal binding of metal ion exit site mutants regarding structures and biochemistry as it could possibly be related to the iron mineralisation. The second mutation site that should be taken into account is Y39, the well-conserved residue in the FOC in a far distance from the di-iron that has been demonstrated to facilitate the iron oxidation of classical ferritin as a single-electron molecular capacitor (Ebrahimi, Hagedoorn & Wilfred R Hagen 2013). It has been always intriguing to understand the structure and biochemistry of the Y39 mutant.

Chapter 5 Iron storage in encapsulin nanocompartments

5.1 Introduction

The encapsulin nanocompartment from *R. rubrum* consists of two proteins, a ferritin-like protein (Rru-EncFtn or Rru-EncFtn) and the encapsulin shell protein Rru-Enc (Rru-Enc). The structure of the ferritin-like protein (EncFtn) of *R. rubrum* has been characterised as an annular decamer (**Section 3.4**), whereas the encapsulin structure is still unknown. The encapsulin family proteins are structurally related to HK97 virus capsid gp5 (Akita et al. 2007). Within the encapsulin family, Rru-Enc shares 33% protein sequence identity with the *T. maritima* encapsulin Tm-Enc (KEGG ID: Tmari_0786, PDB ID: 3DKT) (Sutter et al. 2008); 21% identity with the *M. xanthus* encapsulin Mx-Enc (KEGG ID: MXAN_3556, PDB ID: 4PT2) (McHugh et al. 2014); and 21% identity with *P. furiosus* encapsulin Pf-Enc (KEGG ID: PFC_05175, PDB ID: 2E0Z) (Akita et al. 2007). This chapter described the production of Rru-Enc in *E. coli* and purification via sucrose cushion/gradient ultracentrifugation. To understand the function of the encapsulin nanocompartment in its biochemical context, Rru-EncFtn was coexpressed with Rru-Enc in *E. coli*; purified by sucrose cushion/gradient ultracentrifugation, and further compared with the encapsulin Rru-Enc itself regarding the size distribution using transmission electron microscopy. The iron oxidation and loading capacity were compared in Rru-EncFtn, Rru-Enc and Rru-EncFtn-Enc with apoferritin as a control. Based on assembly, structure and biochemistry data, an iron storage model of encapsulin nanocompartments is proposed.

5.2 Purification of untagged Rru-Enc and Rru-EncFtn-Enc

Rru-Enc was overexpressed in *E. coli* BL21(DE3) using the pACYCDuet-1 vector. The co-expression of Rru-EncFtn and Rru-Enc was achieved by inserting the coding sequences for Rru-EncFtn and Rru-Enc into the two multiple cloning sites on pAYCDuet-1 vector. Coexpressed Rru-EncFtn and Rru-Enc (Rru-EncFtn-Enc) were purified from *E. coli* BL21(DE3) grown in LB, or SeMet medium for SeMet-labelled proteins. The SeMet-labelled proteins can be analysed by ICP-MS to determine the metal to protein ratio; and can also be used in crystallisation with the experimental

phasing method to determine the structure. The purification of the SeMet-labelled proteins is shown below.

5.2.1 Sucrose cushion ultracentrifugation

The Rru-EncFtn-Enc protein complex and Rru-Enc were both purified in GF buffer using 38% (w/v) sucrose cushion ultracentrifugation at 100,000 x *g*, 4°C for 21 hrs. The bottom 3 ml fractions including the pellet were analysed by 15% SDS-PAGE (**Figure 5.2.1**).

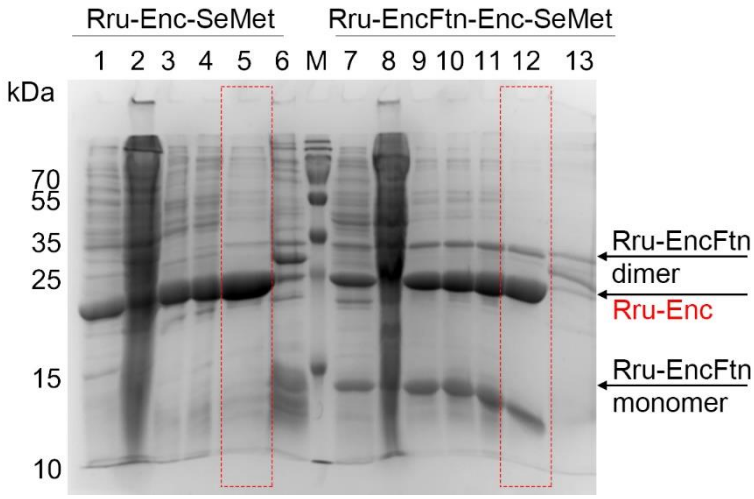


Figure 5.2.1 18% SDS-PAGE analysis of SeMet-labelled encapsulin nanocompartments purified by sucrose cushion ultracentrifugation.

The crude extracts of SeMet-labelled Rru-Enc (Rru-Enc-SeMet) and co-expressed Rru-EncFtn and Rru-Enc (Rru-EncFtn-Enc-SeMet) were treated by 50 µg/ml RNase A for 2 hrs, followed by centrifugation at 48,300 x *g* and 4°C for 20 min; the supernatant of each protein was loaded into lanes 1 and 7 while the pellet was loaded into lanes 2 and 8, respectively. The supernatant of each protein was filtered through 0.45 µm Millipore Syringe Filter before loading to 38% sucrose cushion, followed by ultracentrifugation at 100,000 x *g* and 4°C for 21 hrs. The last 3 ml fractions from each ultracentrifugation tube were loaded to lanes 3–4 and 9–11, respectively. The bottom pellets were fully resuspended in GF buffer and centrifuged to separate the soluble proteins, which were loaded to lanes 5 and 12, respectively. The insoluble part of the bottom pellets was loaded into lanes 6 and 13, respectively. The SDS-PAGE displays a major band closed to 25 kDa, existing in both Rru-Enc-SeMet and Rru-EncFtn-Enc-SeMet fractions, which was identified as Rru-Enc protein via PMF. There are another two major bands around 30 and 15 kDa in Rru-EncFtn-Enc-SeMet, consistent with the dimer and monomer sized Rru-EncFtn. They were subsequently identified as *R. rubrum* EncFtn. Two proteins in lanes 5 and 12 highlighted in red dashed rectangles, respectively, were further purified via sucrose gradient ultracentrifugation.

5.2.2 Sucrose gradient ultracentrifugation

The pelleted fractions from the sucrose cushion were resuspended in GF buffer, followed by removing any insoluble material by centrifugation at 15,000 x *g* and filtering through a 0.22 µm Millipore Syringe Filter. The filtered supernatant was further purified in GF buffer using 10%–50% (w/v) sucrose gradient ultracentrifugation at 100,000 x *g* and 4°C for 17 hrs. Every 0.75 ml fraction was removed from top to bottom of the 38.5 ml centrifuge tubes and the A_{280} and A_{260} of each fraction was determined by using a NanoDrop 2000 spectrophotometer. Fractions of Rru-Enc from 24 ml to the bottom pellet were subjected to SDS-PAGE analysis. Fractions of Rru-EncFtn-Enc from 26 ml to the bottom pellet were analysed by SDS-PAGE as well (**Figure 5.2.2**). The SDS-PAGE gels revealed a major band around 25 kDa in both Rru-Enc and Rru-EncFtn-Enc proteins, corresponding to the monomer size of Rru-Enc. The other two major bands around 15 kDa and 35 kDa in the SDS-PAGE of Rru-EncFtn-Enc correlated with the monomer and dimer sized Rru-EncFtn, suggesting that the coexpression of Rru-EncFtn and Rru-Enc allowed the encapsulation of Rru-EncFtn in the Rru-Enc nanocompartment.

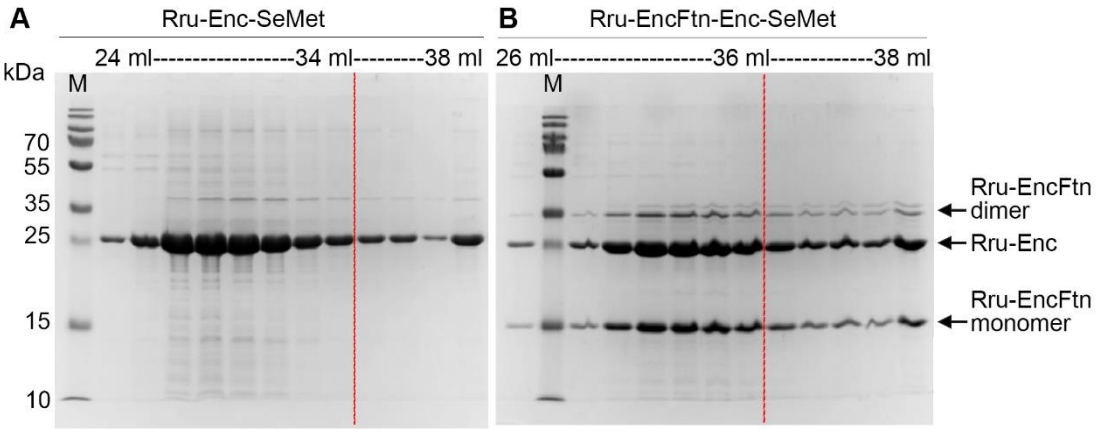


Figure 5.2.2 18% SDS-PAGE analysis of SeMet-labelled encapsulin nanocompartments purified via sucrose gradient ultracentrifugation.

Proteins Rru-Enc and Rru-EncFtn-Enc with SeMet labelling were purified through a 10%–50% (w/v) sucrose gradient by ultracentrifugation. Fractions were removed from each centrifuge tube from top to the bottom. (A) The fractions from 24 ml to 38 ml of Rru-Enc-SeMet and (B) fractions of Rru-EncFtn-Enc-SeMet from 26 ml to 38 ml were subjected to 15% SDS-PAGE analysis. A single major band around 25 kDa shown on the SDS-PAGE of Rru-Enc-SeMet and Rru-EncFtn-Enc-SeMet indicates the monomer of Rru-Enc. Another two major bands around 15 kDa and 35 kDa on the gel of Rru-EncFtn-Enc-SeMet correspond to the monomer and dimer size of Rru-EncFtn.

5.3 TEM visualisation of encapsulin nanocompartments

SeMet-labelled Rru-Enc and Rru-EncFtn-Enc proteins were purified as above and pooled before buffer exchange against 2 litres of GF buffer for two days to remove the sucrose. Proteins were diluted in GF buffer to 0.1 mg/ml prior to visualisation by negative-stain transmission electron microscopy (TEM) experiment (**Figure 5.3.1**). The TEM of negatively stained specimens confirmed that the purified Rru-Enc and Rru-EncFtn-Enc consisted of a population of sphere-like particles of around 24 nm in diameter, consistent with the size of the encapsulin (KEGG ID: Tmari_0786, PDB ID: 3DKT) from *T. maritima* as a 60-mer encapsulin nanocompartment (Sutter et al. 2008). Micrographs of purified Rru-EncFtn-Enc revealed 24 nm nanocompartments with regions in the centre that excluded stain, consistent with the presence of the Rru-EncFtn within the encapsulin shell. The size distribution of each protein complex was analysed in the software GraphPad Prism 6.0 by measuring the diameter of each nanoparticle in three different directions using ImageJ (Abramoff et al. 2004). The similar size of Rru-Enc and Rru-EncFtn-Enc indicated that the packaging of Rru-EncFtn did not alter the size of Rru-Enc significantly.

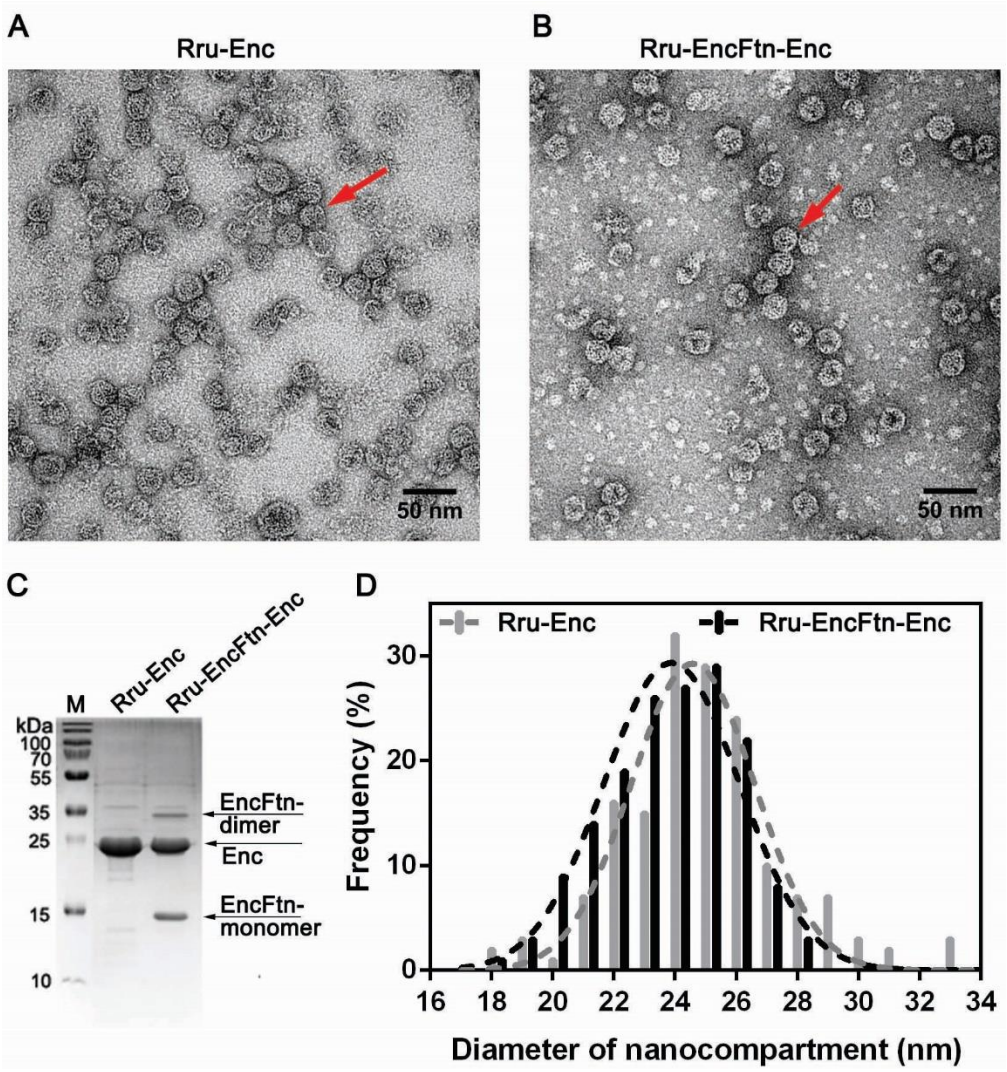


Figure 5.3.1 TEM visualisation of *R. rubrum* encapsulin nanocompartments and their size distribution.

Red arrows highlight the nanocompartments in TEM images of SeMet labelled (A) Rru-Enc and (B) Rru-EncFtn-Enc stained by 2% uranyl acetate. All samples were imaged at 143,000 *x* magnification; the scale bar corresponds to 50 nm. (C) 15% SDS-PAGE analysis of Rru-Enc and Rru-EncFtn-Enc. (D) Histogram of size distribution of Rru-Enc and Rru-EncFtn-Enc particles. A Gaussian curve was fitted to the data by non-linear least squares regression to obtain a mean diameter of 24.6 nm with a standard deviation of 2.0 nm for Rru-Enc (grey) and a mean value of 23.9 nm with a standard deviation of 2.2 nm for Rru-EncFtn-Enc (black).

5.4 Crystallisation of Rru-Enc

Both native and SeMet labelled Rru-Enc at 10 mg/ml was subjected to crystallisation trials using MRC 96 well crystallisation plates and the sitting drop vapour diffusion technique. An Art Robbins Gryphon crystallisation robot was used with commercial screen kits including PEG/Ion HT from Hampton Research, JCSG-plus HT-96 and Structure Screen 1 & 2 HT-96 and MIDAS from Molecular Dimensions. Crystals of

Rru-Enc has grown in a bi-pyramidal morphology in a number of different crystallisations conditions. **Figure 5.4.2 A–B** shows a number of representative crystals of Rru-Enc. Crystals of both the native and SeMet labelled Rru-Enc did not diffract with a d_{max} better than 10 Å. The poor diffraction quality may be a result of impurities carried over from the sucrose cushion/gradient ultracentrifugation. Therefore, a pET-28a(+) based Rru-Enc construct with a C-terminal his-tag (Rru-Enc-his) was overexpressed in *E. coli* for structural determination. This protein was purified through IMAC and SEC (**Figure 5.4.1**). However, Rru-Enc-his readily precipitated during dialysis or concentration by centrifugal ultrafiltration. Although much purer protein was obtained using this construct it did not improve the diffraction resolution (**Figure 5.4.2 C**).

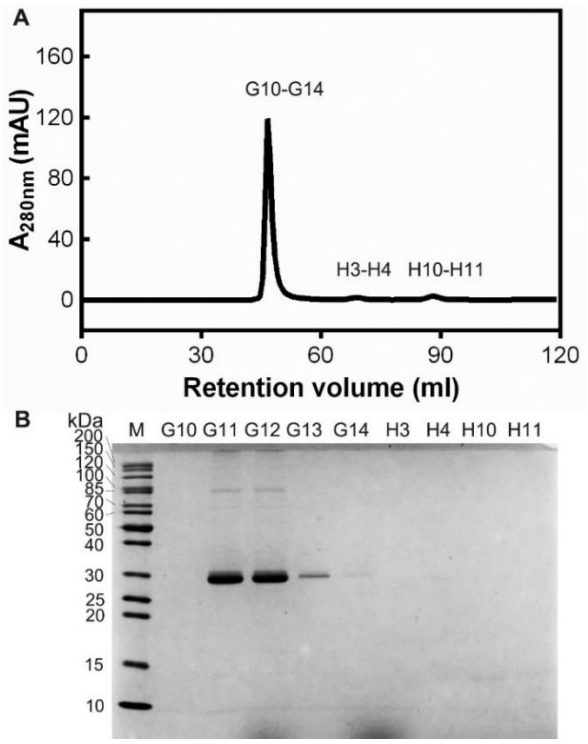


Figure 5.4.1 Size exclusion chromatography purification of Rru-Enc-his using a HiLoad 16/600 Superdex 200 pg column.

(A) The chromatogram of SEC purification of Rru-Enc-his shows a major peak (fractions G10–G14) with a retention volume of 46 ml, which is close to the void volume of the column. Two small peaks are shown at 68 ml (fractions H3–H4) and 88 ml (fractions H10–H11). (B) Fractions labelled in (A) were subjected to 15% SDS-PAGE. There is a major band around 30 kDa and some contaminant bands around 85 kDa. The major band correlating with the size of Rru-Enc-his, was later identified as Rru-Enc by PMF. Fractions G11 and G12 were pooled for crystallisation experiments.

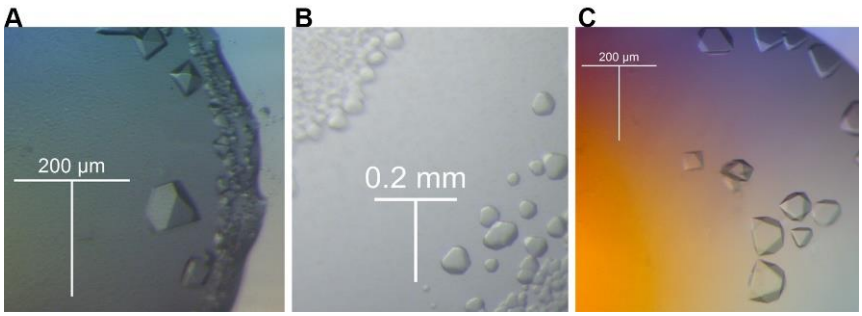


Figure 5.4.2 Images of crystals of Rru-Enc, Rru-Enc-SeMet and Rru-Enc-his.

(A) Rru-Enc purified from LB medium by sucrose cushion/gradient ultracentrifugation crystallised in the E11 condition of the PEG/Ion kit containing 4% (v/v) tacsimate pH 5.0 and 12% (w/v) PEG 3350. (B) The SeMet-labelled Rru-Enc purified from SeMet medium using sucrose cushion/gradient ultracentrifugation crystallised in the condition containing 0.25 M KNO_3 and 20% (w/v) PEG 3350. (C) The Rru-Enc with C-terminal his-tag (Rru-Enc-his) purified through IMAC and SEC was grown into bi-pyramidal crystals in the A10 position of the MIDAS screen kit, which contains 24% (w/v) polyvinyl pyrrolidone K15 and 0.1 M sodium sulphate. The scale bar indicates 200 µm in each image.

5.5 Ferroxidase assay of EncFtn and encapsulin

It is generally accepted that iron storage in ferritin cages involves Fe(II) oxidation to Fe(III) and the subsequent formation of iron oxide-hydroxide and iron phosphate minerals. These reactions are catalysed by the ferroxidase centre of the ferritin structure. In light of the identification of an iron-loaded FOC in the crystal structure of EncFtn and the native mass spectrometry data, peroxidase and ferroxidase assays were performed to demonstrate the catalytic activity of this protein. In addition, equine apoferritin, an example of a classical ferritin enzyme, was also assayed as a positive control. Unlike the DPS family of ferritin-like proteins, EncFtn showed no peroxidase activity when assayed with the substrate *ortho*-phenylenediamine (Pesek et al. 2011).

5.5.1 Iron oxidation by encapsulin nanocompartments

To measure the ferroxidase activity of Rru-EncFtnSH, Fe(II) was added to the protein in GF buffer at 25°C and progress curves for the formation of Fe(III) minerals were recorded for 1800 s at 315 nm. Fe(II) was added to the reactions at 20 and 100 µM final concentration (2 and 10 times molar ratio of Fe(II)/FOC). The concentration of Rru-EncFtnSH was determined using a BCA microplate assay prior to the Fe(II) loading. Each solution used in these assays was prepared anaerobically including fresh $Fe(NH_4)_2(SO_4)_2$ in 0.1% (v/v) HCl, GF buffer and diluted protein samples. Decamer fractions of purified Rru-EncFtnSH were used in the assays to avoid confounding

effects from protein assembly in the presence of Fe(II). Since Rru-EncFtn is encapsulated in Rru-Enc in the native bacterium, *R. rubrum*, the ferroxidase activity of both recombinant Rru-Enc and coexpressed Rru-EncFtn and Rru-Enc (Rru-EncFtn-7374) was also measured. As the Rru-Enc structure was not available and it presumably resembles the *T. maritima* encapsulin, Tmari_0786, based on the TEM observation as well, the recombinant Rru-Enc was very likely to be a 60-meric T=1 icosahedron with 12 pentamers. Therefore, the Rru-Enc nanocompartment can sequester no more than twelve decameric Rru-EncFtn, leading to a maximum molar ratio 2:1 of Rru-EncFtn: Rru-Enc. The ribosome binding site (RBS) calculation of the mRNA sequence of pACYCDuet-1-Rru-EncFtn-Rru-Enc estimated that the 'ATG' translation rate of recombinant Rru-EncFtn was twice as high as the counterpart of Rru-Enc, suggesting a 2:1 molar ratio of Rru-EncFtn: Rru-Enc in the co-expression system in *E. coli* BL21(DE3) (Espah Borujeni et al. 2014; Salis et al. 2009). Therefore, 10 µM Rru-EncFtn-Enc was used in this ferroxidase assay based on the estimation that the recombinant Rru-Enc was fully occupied by Rru-EncFtn (**Figure 5.7.4**). It is worth mentioning that the RBS calculation was based on a biophysical model developed by Salis lab that predicts translation initiation rate for each start codon in an mRNA sequence. Thus it is likely that the RBS calculation might be inaccurate to some extent. In this case, obitrap mass spectrometry can be very instrumental in solving the stoichiometry of encapsulin nanocompartment proteins (Snijder et al. 2014).

To explore if Rru-Enc contributed enzymatic activity towards iron storage in encapsulin nanocompartments, Fe(II) was also added to Rru-Enc at a proportional protein concentration to the Rru-EncFtn-Enc complex. Horse spleen apo-ferritin (apo-HsFtn) purchased from Sigma, prepared as iron-free sample was also used in this ferroxidase assay to demonstrate the validity of this method and as a positive control for ferroxidase activity. Since the mammalian ferritin monomer has a single ferroxidase centre (FOC) in each of the four-helix bundle monomers, a monomer of apo-HsFtn has an equivalent number of FOCs to a dimer of Rru-EncFtnSH. Therefore, the concentration of apo-HsFtn used in these experiments was half of the concentration of Rru-EncFtnSH. Finally, iron background oxidation was monitored by mixing Fe(II) with GF buffer in air. As the ferroxidase assays in panels A and B (**Figure 5.5.1**) were

performed on different days, iron background oxidation was also measured on each day to reduce the variation of the assays.

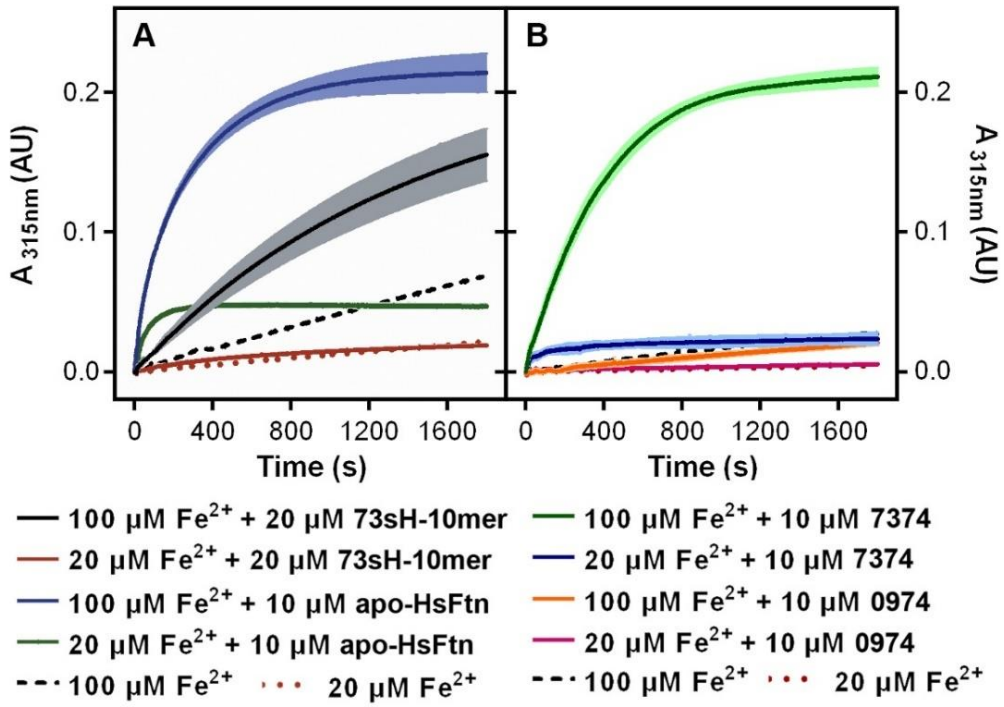


Figure 5.5.1 Progress curve of Fe(II) oxidation catalysed by encapsulin nanocompartments.

(A) 20 μM Rru-EncFtnSH decamer fraction (73sH-10mer, 10 μM FOC) or 10 μM horse spleen apo-ferritin (apo-HsFtn, 10 μM FOC) was incubated with 20 or 100 μM Fe(II) in GF buffer (pH 8) at 25°C for 1800 s with absorbance at 315 nm (A_{315}) monitored. Fe background oxidation was detected by mixing 100 or 20 μM Fe(II) with GF buffer. (B). 10 μM encapsulin nanocompartment with co-expressed Rru-EncFtn and Rru-Enc (7374) or encapsulin Rru-Enc (0974) was mixed with 20 or 100 μM Fe(II) in GF buffer. The results are averaged from three technical repeats with error bars indicating the standard deviation from the mean shown.

The ferroxidase assay demonstrated that Rru-EncFtnSH was catalytically active to oxidise Fe(II) to Fe(III). When compared to apoferritin, Rru-EncFtnSH oxidised Fe(II) at a slower rate and the reaction did not appear to run to completion over the 1800 seconds of the experiment. Addition of higher quantities of iron resulted in the formation of a yellow/brown precipitate at the end of the reaction, which could be the insoluble iron minerals formed on the surface of Rru-EncFtnSH, and subsequent precipitation of the Rru-EncFtnSH. The Rru-Enc encapsulin protein did not show iron oxidation ability above the background Fe(II) oxidation. The ferroxidase activity of Rru-EncFtnSH is retained and enhanced when encapsulated by Rru-Enc as a co-expressed encapsulin nanocompartment, with this sample showing a faster iron

oxidation rate than Rru-EncFtnSH on its own and a comparable ferroxidase activity with apo-HsFtn (10 Fe(II) / 2 FOC).

5.5.2 Iron oxidation by Rru-EncFtnSH FOC mutants

To investigate the biochemical role played by the metal binding residues in the FOC of Rru-EncFtnSH, the ferroxidase activity for a number of FOC mutants was investigated, including E32A, E62A and H65A (**Figure 5.5.2**). Compared to the wild-type, all mutants demonstrated substantially reduced ferroxidase activities, of which the E62A mutant almost diminished all Fe(II) oxidation capability. These data, coupled with the mutant structures (**Section 4.4**), confirmed the central role of ferroxidase centre in achieving the ferroxidase activity of the EncFtn protein.

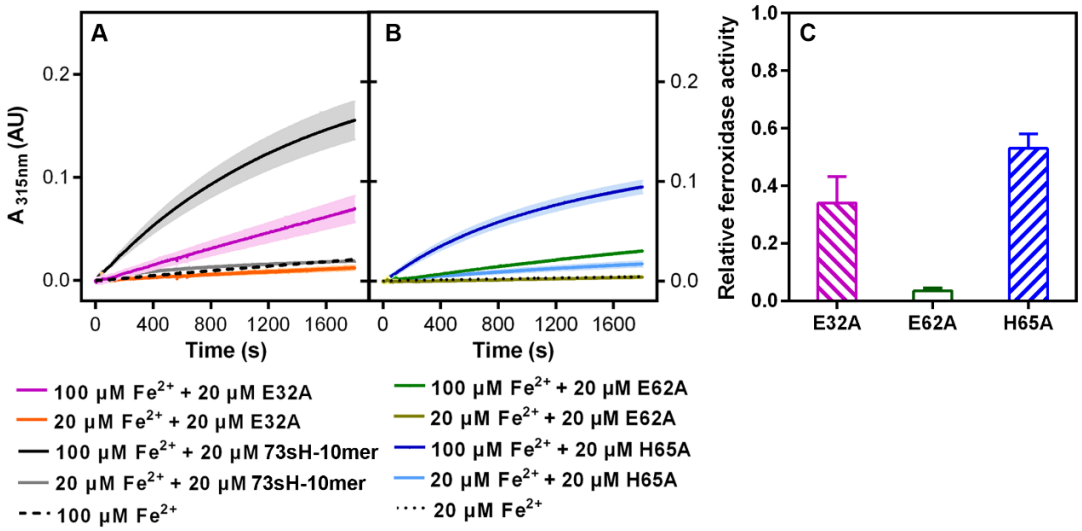


Figure 5.5.2 Spectroscopic evidence for Fe(II) oxidation by FOC mutants of Rru-EncFtnSH.

FOC mutants of Rru-EncFtnSH including E32A, E62A and H65A at 20 μM were mixed with 100 μM or 20 μM Fe(II) in GF buffer (pH 8) at 25°C and the progress curve at A315 was monitored for each sample for 1800 seconds. A negative control of Fe(II) background oxidation, and positive control of wild-type decameric Rru-EncFtnSH (73sH-10mer) catalysis are both shown for reference. The experiments were performed as three technical repeats with the error bars indicating the standard deviation of the mean for each sample. (A) E32A *vs* wild-type Rru-EncFtnSH; (B) E62A *vs* H65A. (C) The relative ferroxidase activity of mutants is plotted as a proportion of the activity of the wild-type protein using the endpoint measurement of A_{315}. The error bars represent the calculated standard deviations of three technical repeats. The FOC mutants showed reduced ferroxidase activity to varied extents, with the E62A mutant significantly abrogated the ferroxidase activity.

5.6 Iron storage capacity of encapsulin nanocompartments

Since the iron oxidation catalysed by Rru-EncFtnSH did not fully finish in the 1800 seconds recorded for the progress curves, the final iron loading capacity of each protein was determined after 3-hour-incubation of protein with different amounts of Fe(II). Before iron determination by ferrozine assay, the free iron ions in samples were be removed by desalting with a pre-equilibrated Zeba™ spin desalting column (7K MWCO, Thermo Scientific).

5.6.1 Standard curve of ferrozine assay

$FeCl_3$ was used in the ferrozine assay as the iron standard. The ferrozine assay involves three steps: iron release from proteins using a mixture of 0.7 M HCl and 2.25% (w/v) $KMnO_4$ at 60°C for 2 hrs; iron chelation by a mixture of 6.5 mM ferrozine, 6.5 mM neocuproine, 2.5 M buffer (ammonium acetate) and reducing agent (1 M ascorbic acid) in room temperature for 30 min; and detection by measuring the absorbance at 562 nm for the detection of the -ferrozine complex (Riemer et al. 2004). The standard curve for the ferrozine assay was plotted using Fe(III) concentration (μM) as the X-axis and A_{562nm} as the Y-axis (**Figure 5.6.1**). The iron content in proteins was detected in the same procedure, with the standard curve used to estimate the iron content of the sample.

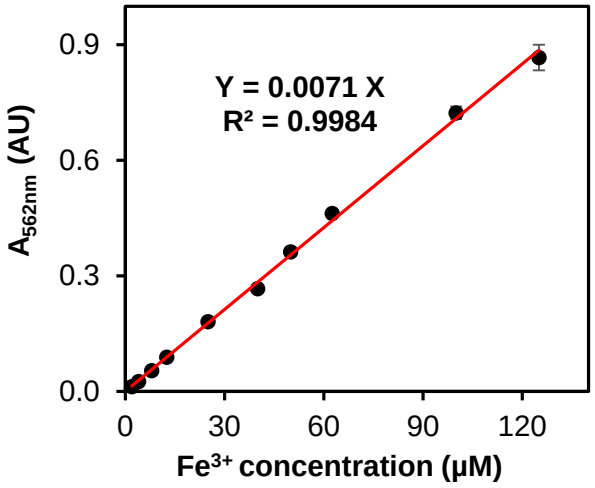


Figure 5.6.1 Ferrozine standard curve.

$FeCl_3$ standards of concentrations (black sold circles) ranging from 0 to 120 μM are applied in the ferrozine assay. The corresponding A_{562} is recorded to plot against each $FeCl_3$ concentration and fitted with a red linear regression line with the equation and R-squared correlation shown. This experiment has three technical repeats with the error bars indicated by standard deviation.

5.6.2 Standard curve of BCA microplate assay

The concentration of proteins used in these assays was determined using the Pierce™ BCA Protein Assay Kit based on the manufacturer's instructions (Thermo Fisher Scientific). The bovine serum albumin (BSA) standard was used in the BCA microplate assay and a standard curve was plotted using BSA concentration as the X-axis and absorbance at 562 nm as the Y-axis.

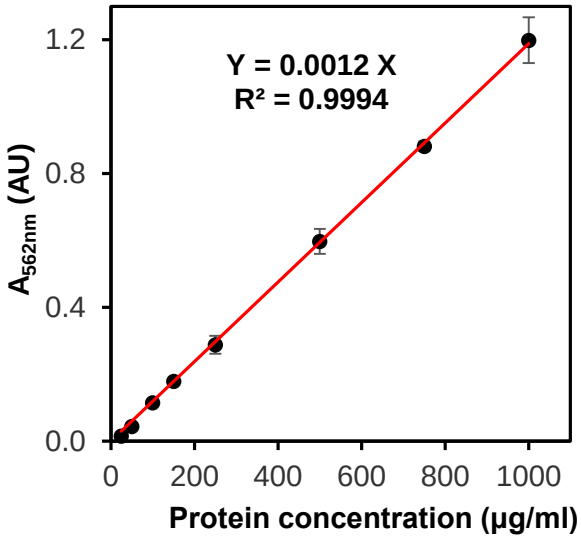


Figure 5.6.2 Standard curve of BCA microplate assay.

25 µl BSA in GF buffer at 25, 50, 100, 150, 250, 500, 750 and 1000 µg/ml concentrations were mixed with 200 µl of freshly prepared working solution in a 96-well plate with three technical repeats. The absorbance at 562 nm was measured after incubation at 37°C for 30 min. Each BSA standard at different concentrations (black solid circles) is fitted with a red linear regression line with the equation and R-squared correlation shown.

5.6.3 Iron loading capacity of encapsulin nanocompartments *vs* EncFtn

Protein samples at 8.5 µM were mixed with various concentrations of acidic Fe(II) at room temperature for three hours before removal of protein and mineral precipitation by centrifugation and desalting with a centrifugal desalting column. The iron content and protein concentration were determined on each desalted sample by the ferrozine assay and BCA microplate assay, respectively (**Figure 5.6.3 A–D**).

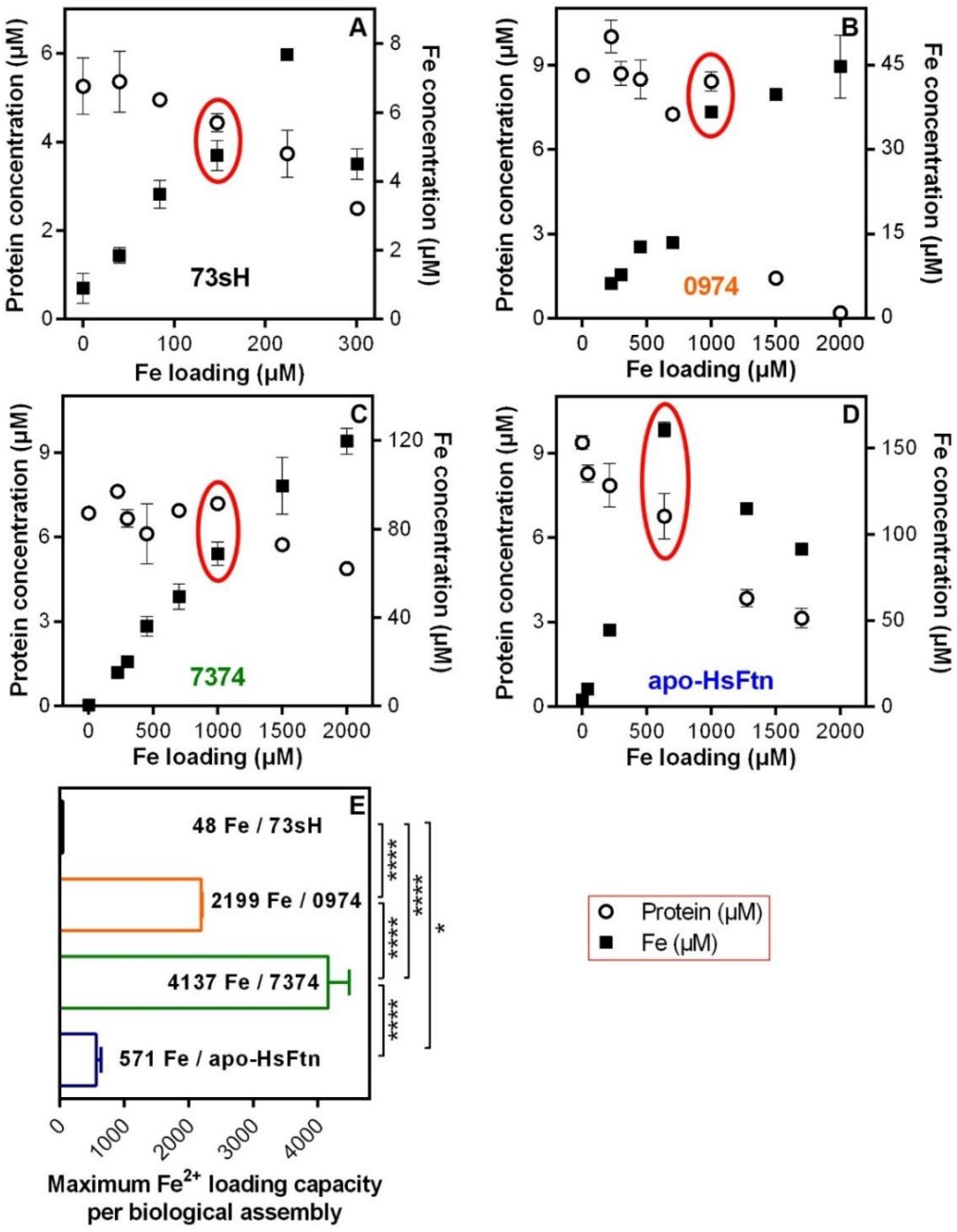


Figure 5.6.3 Comparison of iron loading capacities of EncFtn, encapsulin and ferritin.

Fe(II) at various concentrations were mixed with 8.5 μM (A) decameric Rru-EncFtnSH (73sH), (B) Rru-Enc (0974), (C) and co-expressed Rru-EncFtn and Rru-Enc (7374), or (D) horse spleen apo-ferritin (apo-HsFtn) and incubated at room temperature for three hours before desalting. Iron content (black solid squares) and protein concentration (black empty circles) are plotted for each sample. (E) The maximum iron loading capacity of each protein as a biological assembly was calculated based on the protein-iron sample (highlighted by red ellipses) with the highest iron concentration and the least protein loss. Error bars indicate the standard deviation from the mean of three technical repeats. The data were statistically compared by one-way ANOVA analysis (GraphPad Prism 6.0 software). Asterisks indicate significant differences between pairs of proteins regarding the maximum Fe loading capacity. Four asterisks mean extremely significant with $p < 0.0001$; while one asterisk indicates the difference between apo-HsFtn and Rru-EncFtnSH is significant ($p < 0.05$).

Protein samples displayed a limited iron storage capacity to different degrees since yellow/brown precipitate started to form when too much Fe(II) was loaded. The precipitate could arise from background iron oxidation by air when the ferroxidase was saturated; or the denatured protein when it was overloaded by iron. Rru-EncFtnSH-Fe mixture started to precipitate when more than 150 μM Fe(II) was present. Both Rru-Enc and Rru-EncFtn-Enc became instable in the presence of over 1 mM Fe(II); but Rru-Enc concentration reduced more sharply than the coexpressed protein complex. Apoferritin showed a certain stability when less than 700 μM Fe(II) was loaded. Therefore, the protein stability in terms of the Fe(II) tolerance was shown in an order: Rru-EncFtn-Enc > Rru-Enc > apo-HsFtn > Rru-EncFtnSH. The iron loading of the protein samples with significant precipitation has not been taken into account, since it was not possible to directly measure the distribution of iron and protein in both the supernatant and precipitation. Therefore, the maximum iron concentration with the least protein loss was used to derive the maximum iron loading capacity of the samples in this experiment.

The maximum iron loading capacity of the protein samples was calculated as the quantity of iron per biological assembly, which corresponded to ten subunits of Rru-EncFtnSH, 60 subunits of Rru-Enc, 60 subunits of Rru-EncFtn-Enc (2:1 of Rru-EncFtn: Rru-Enc) and 24 subunits of apo-HsFtn. Therefore, the Rru-EncFtnSH decamer bound a maximum of around 47 iron ions before excess iron induced the protein precipitation. The encapsulin protein Rru-Enc could sequester about 2,199 iron ions before significant protein loss occurred, and the Rru-EncFtn-Enc sequestered about 4,137 iron ions. This latter result was significantly more than the apoferritin binding ability determined in this assay, which housed approximately 570 iron ions (**Figure 5.6.3 E**).

The Rru-EncFtnSH could bind iron ions at the FOC, metal ion entry/exit sites as well as the surface acidic residues, which are likely to contribute to the 47 iron ions detected in this experiment. While the isolated encapsulin Rru-Enc did not display any ferroxidase activity, it bound around 2,000 iron ions in this assay. This implies that the encapsulin nanocompartment shell could bind a significant amount of iron ions on its outer and inner surfaces. The Rru-EncFtn-Enc complex was able to bind seven times more iron than the apoferritin. The ferroxidase assay and iron-binding measurements

highlight the different roles that Rru-EncFtn and Rru-Enc play in the oxidation and storage of iron ions in encapsulin nanocompartments: Fe(II) is oxidised by Rru-EncFtn into Fe(III) and its mineralisation product is then stored in the cavity of Rru-Enc. It did not escape our notice that apoferritin in this experiment did not display comparable iron loading ability to the literature, which cited an iron binding capacity of between 1,000 and 4,000 iron ions (Le Brun et al. 2010). This is likely a consequence of the iron loading method used in this experiment; other experimental approaches used a step-wise addition of iron to load ferritin, which might allow an increased loading capacity to be measured, due to lower levels of protein damage and loss through precipitation. This also implies that the bacterial naoncompartments might have not been fully occupied by iron. With experimental optimisation and progressive iron loading, it may be possible to record higher levels of iron storage for these complexes.

5.6.4 Visualisation of iron-loaded proteins by TEM

Rru-EncFtnSH was too small to be visualised under the TEM. To visualise the products of iron loading in the samples and to determine whether the EncFtn and encapsulin form amorphous iron precipitates, or ordered iron nanoparticles, the four protein samples including Rru-EncFtnSH, Rru-Enc, Rru-EncFtn-Enc, and horse spleen apoferritin were loaded with based on the maximum loading ratio (**Figure 5.6.3**) for 1 hour at room temperature before subjecting them to transmission electron microscopy (TEM) analysis with or without stain by uranyl acetate (**Figure 5.6.4**). The Rru-EncFtnSH-Fe mixture showed the formation of large, irregular electron-dense precipitates (**Figure 5.6.4 A**). A similar distribution of particles was observed after addition of Fe(II) to the encapsulin protein (**Figure 5.6.4 B**). In contrast, addition of Fe(II) to the Rru-EncFtn-Enc nanocompartment resulted in small, highly regular, electron dense particles of approximately 5 nm in diameter (**Figure 5.6.4 C**); this was interpreted as controlled mineralisation of iron within the nanocompartment. Addition of Fe(II) to apoferritin resulted in a mixture of large particles and small (~2 nm) particles consistent with partial mineralisation by the ferritin and some background oxidation of the iron (**Figure 5.6.4 D**). Negative stain TEM of these samples revealed that upon addition of iron, the Rru-EncFtnSH protein showed significant aggregation (**Figure 5.6.4 F**); while the Rru-Enc, Rru-EncFtn-Enc system, and apoferritin were

present as distinct nanocompartments without significant protein aggregation (**Figure 5.6.4 G–I**).

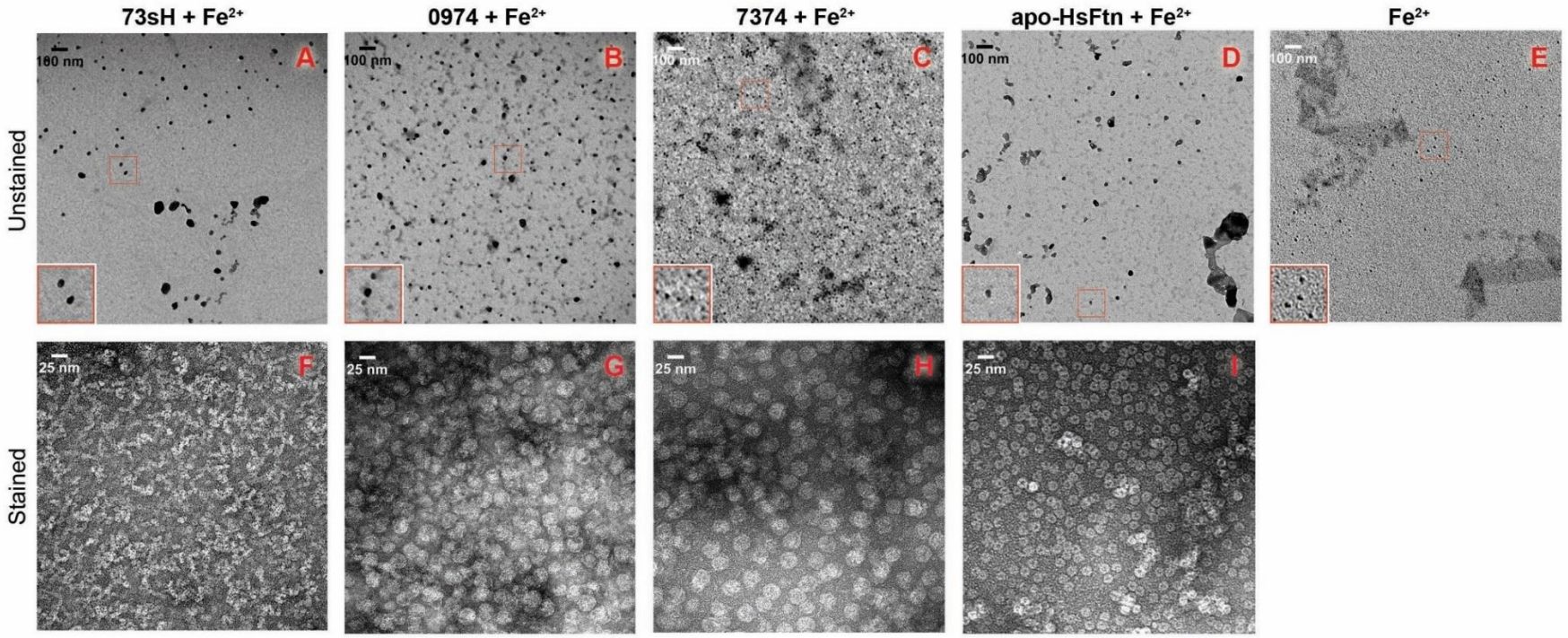


Figure 5.6.4 TEM visualisation of iron-loaded EncFtn, encapsulin, and ferritin protein samples.

Decameric Rru-EncFtnSH (73sH), encapsulin Rru-Enc (0974), co-expressed Rru-EncFtn and Rru-Enc (7374), and horse spleen apoferritin (apo-HsFtn), at 8.5 µM, were mixed with 147 µM, 1 mM, 1 mM and 215 µM acidic $Fe(NH_4)_2(SO_4)_2$ respectively. Protein mixtures were incubated at room temperature for 1 hour prior to TEM analysis with or without uranyl acetate stain. (A-D) Unstained 73sH, 0974, 7374, apo-HsFtn loaded with Fe(II), respectively, with 35,000 *x* magnification and scale bars indicating 100 nm. (E) Protein-free sample as a control. (F-I) Uranyl acetate stained 73sH, 0974, 7374, apo-HsFtn loaded with Fe(II), respectively, with 140,000 *x* magnification and scale bars indicating 25 nm. The insets are bigger views corresponding to the small red square in each image.

5.7 The iron storage in encapsulin nanocompartments

5.7.1 Comparison of encapsulins

As EncFtn is natively enclosed inside an encapsulin nanocompartment, the surface pores in encapsulin are therefore of great importance for the movement of substrates for the EncFtn across the nanocompartment shell. Herein, the structures of three encapsulins including Tm-Enc from *T. maritima* (KEGG ID: Tmari_0786, PDB ID: 3DKT, resolution: 3.1 Å) (Sutter et al. 2008), Pf-Enc from *P. furious* (KEGG ID: PFC_05175, PDB ID: 2E0Z, resolution: 3.6 Å) (Akita et al. 2007) and Mx-Enc from *M. xanthus* (KEGG ID: MXAN_3556, PDB ID: 4PT2, resolution: 4.6 Å) (McHugh et al. 2014) were analysed for the possibility of molecule trafficking through their surface pores.

Tm-Enc is a 60-meric icosahedron with 12 pentamers assembling in $T = 1$ symmetry. Both Pf-Enc and Mx-Enc form a 180-meric icosahedron, with 12 pentamers and 20 copies of hexamers assembled in $T = 3$ symmetry. The *R. rubrum* encapsulin Rr-Enc (KEGG ID: Rru-Enc) shares 33% amino acid sequence identity with Tm-Enc (Sutter et al. 2008), and 21% identity with Pf-Enc (Akita et al. 2007) and Mx-Enc (McHugh et al. 2014) (**Figure 5.7.1**). Among the three encapsulin structures reported, only Tm-Enc shows a similar size to Rru-Enc as determined by TEM (**Figure 5.3.1**), indicating that Rru-Enc is very likely to be a 60-meric $T = 1$ icosahedron assembled as 12 pentamers. The multiple sequence alignment of these encapsulin proteins highlights the 5-fold or quasi-6-fold symmetry pore, which may act as the entry and exit sites of iron ions and other substrate molecules (**Figure 5.7.1**).

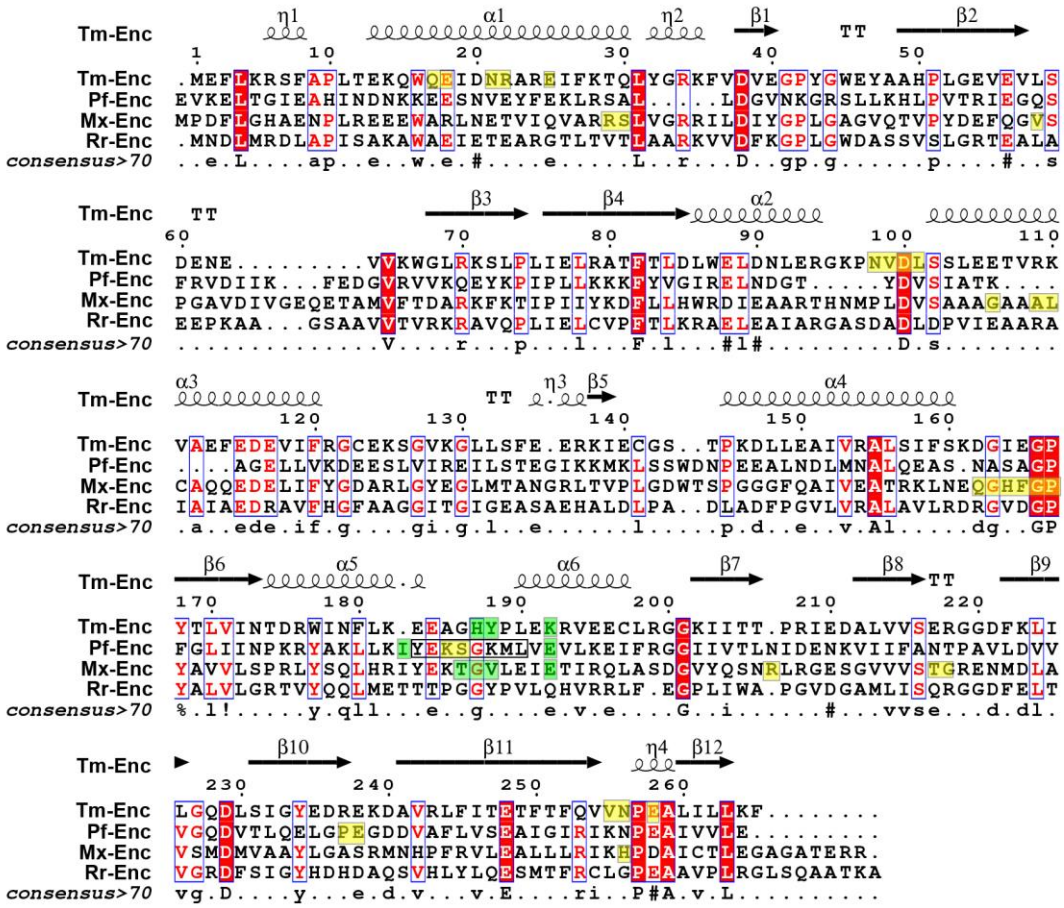


Figure 5.7.1 Multiple sequence alignment of encapsulin proteins.

Encapsulin protein sequences were aligned using Clustal Omega (Sievers et al. 2011) including *T. maritima* encapsulin Tm-Enc (PDB ID: 3DKT) (Sutter et al. 2008), *P. furiosus* encapsulin Pf-Enc (PDB ID: 2E0Z) (Akita et al. 2007), 4PT2 (PDB ID) from *M. xanthus* encapsulin Mx-Enc (PDB ID: 4PT2) (McHugh et al. 2014) as well as the *R. rubrum* encapsulin Rr-Enc (KEGG ID: Rru-Enc). The alignment was presented using Espript 3.0 (Robert & Gouet 2014). The secondary structure of Tm-Enc is shown on top of the alignment. Consensus sequence over 70% identity is displayed at the bottom of the alignment. Dots indicate gaps in the alignment. Protein residues around the 5-fold symmetry pore of encapsulin pentamers are highlighted in green. Residues highlighted in yellow are on encapsulin 6-fold symmetry, or surface pores. The central channels and surface pores are potential channels for molecule transportation (i.e. iron).

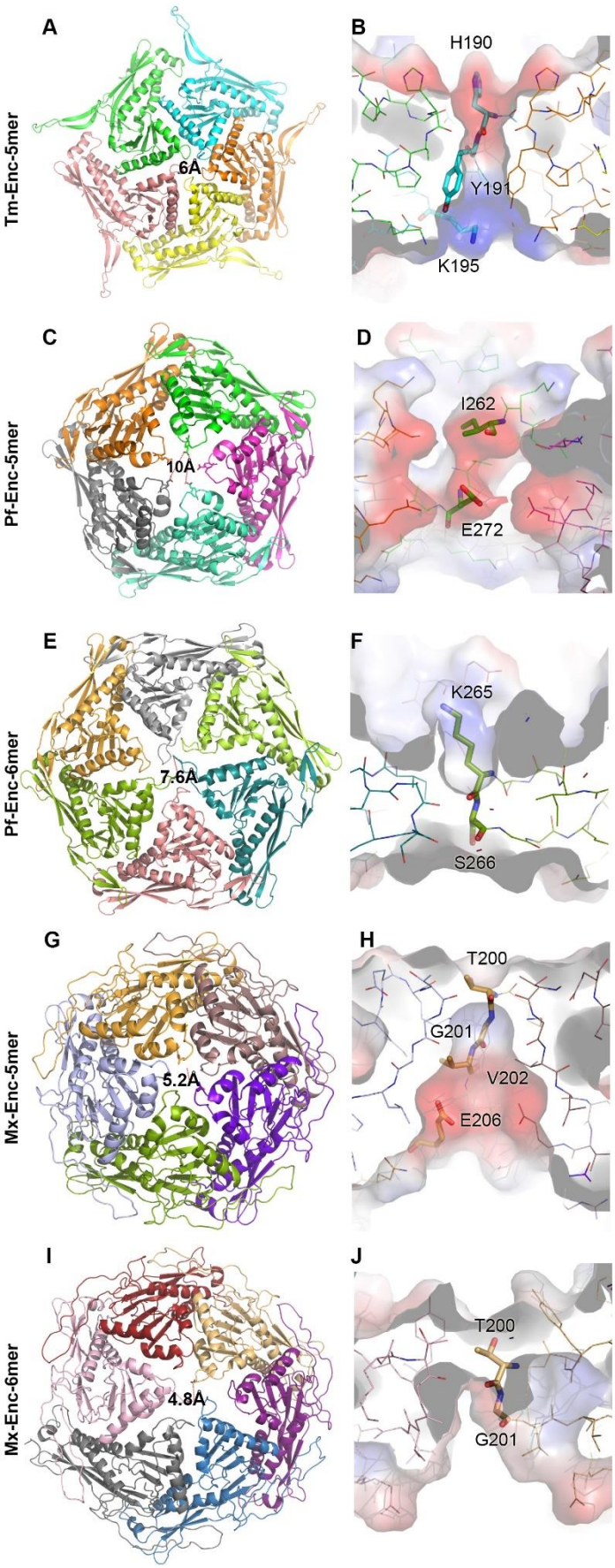


Figure 5.7.2 Encapsulin residues of the 5-fold or 6-fold symmetry pore.

Encapsulins from (A/B) T. maritima (Tm-Enc), (C–F) P. furiosus (Pf-Enc) and (G–J) M. xanthus (Mx-Enc) are shown in their pentamer and/or hexamer assembly, which are presented in cartoon mode and coloured by chain. The central pore of the encapsulin pentamer or hexamer is labelled with the diameter (Å). (B/D/F/H/J) a cut-away view of the encapsulin 5-fold or 6-fold symmetry channels are shown in electrostatic potential surface mode with pore residues highlighted as sticks. (A/B) The 5-fold symmetry channel of Tm-Enc pentamer is 6 Å wide and contains residues H190, Y191 and K195. (C/D) The 5-fold symmetry channel of Pf-Enc pentamer is 10 Å wide and contains residues I262 and E272. (E/F) The 6-fold symmetry channel of Pf-Enc hexamer is 7.6 Å wide and includes residues K265 and S266. (G/H) The 5-fold symmetry channel of Mx-Enc pentamer is 5.2 Å wide and contains residues T200, G201, V202 and E206. (I/J) The 6-fold symmetry channel of Mx-Enc is 4.8 Å wide and consists of residues T200 and G201.

The Mx-Enc has a pore size of around 5 Å in the pentamer and hexamer central channels; this contrasts with the 6 Å pore of the of Tm-Enc pentamer and 7.6 Å of the Pf-Enc hexamer (**Figure 5.7.2 A/E/G/I**). All of the other surface pores have a diameter of around 5–6 Å at the narrowest region. Notably, residues 263–270 'YEKSGKML' of Pf-Enc are not shown in the 5-fold symmetry channel due to the missing density in the structure, which suggests that the 10 Å pore size may be an overestimation (**Figure 5.7.2 C/D**). This observation also indicates the greater levels of flexibility in the 5-fold symmetry pore than the 6-fold symmetry pore in the Pf-Enc structure. It is possible that the central channels of pentamer and hexamer of encapsulins may display substrate specificity and transfer molecules at different rates according to the need for iron storage or mobilisation.

The 5-fold symmetry channel of encapsulins appears like an upside-down funnel, in which the residues are relatively less conserved between the different encapsulin proteins. The 5-fold and 6-fold symmetry pores of 180-meric encapsulins are very likely to share the same residues, as suggested by T200 and G201 in the symmetry pores of Mx-Enc pentamer and hexamer (**Figure 5.7.2 H/J**). However, the T200 carboxylate oxygen points towards the central channel of Mx-Enc pentamer while the T200 side chain methyl group points towards the hexamer central channel. This suggests that the 180-meric encapsulin proteins utilise different orientations of the pentamer and hexamer central pore residues to allow the entry/exit of various molecules (**Figure 5.7.2 C–J**). The K265 in the Pf-Enc hexamer central channel suggests that the 6-fold symmetry pore may transport other molecules such as water, or phosphate ions instead of iron (**Figure 5.7.2 F**). The methyl side chain of T200 and G201 of the Mx-Enc hexamer central channel are also unlikely to coordinate iron.

According to the sequence alignment of the encapsulin proteins, the 5-fold symmetry pore of Rr-Enc consists of the residues 'TTPGGYPVLQ' (**Figure 5.7.1**), in which the underscored GY---Q correspond to the HY---K labelled in Tm-Enc, the GV---E of Pf-Enc, and GK---E of Mx-Enc (**Figure 5.7.2 B**). In the Tm-Enc 5-fold symmetry pore, the negatively-charged outer surface arises from the backbone carbonyl oxygen of H190 while the side chain nitrogen of K195 contributes to the positively-charged

interior surface (**Figure 5.7.2 A/B**); this electrostatic gradient would allow the entry of iron into the *T. maritima* encapsulin, but not its release by this route.

Taken together, the 5-fold symmetry pore is very likely to be the main Fe(II) entry site in both the 60-meric and 180-meric encapsulins. The 6-fold symmetry pore of 180-meric encapsulins could transport other molecules. The presence of other encapsulin surface pores does not rule out the possibility that they could also be involved in molecule trafficking across the encapsulin shell. Future experiment will focus on understanding the structure of Rr-Enc and the role that the residues identified in these structural analysis play in substrate transport across the encapsulin shell.

5.7.2 Iron storage model of 60-meric encapsulin nanocompartments

Since Rru-Enc crystals failed to diffract to higher than 10 Å resolution, the protein structure remains elusive. To understand how iron storage can occur in encapsulin nanocompartments, Rru-EncFtnSH as a representative of the EncFtn family was docked inside the Tm-Enc crystal structure, and a model for iron metabolism by the nanocompartment is proposed based on this analysis (**Figure 5.7.3**). Given that the decameric EncFtn associates with the inner shell of the encapsulin through its C-terminal localisation sequence, EncFtn decamer is most likely to dock into the pentameric vertex of the shell rather than hexamer due to the arrangement of EncFtn structure. It is possible that the FOC dimer of EncFtn, or even monomer, may exist inside the encapsulin as iron nucleation sites (McHugh et al. 2014).

Although the C-terminal localisation sequence was missing in the truncated Rru-EncFtnSH structure, the positions of the C-termini of the monomers in the Rru-EncFtnSH structure and the localisation sequences found in Tm-Enc structure are spatially consistent. Thus, it appears that the EncFtn decamer is the physiological state of this protein. Docking the decamer in a consistent position with the localisation sequences in the Tm-Enc, places the central ring of EncFtn directly above the pore at the five-fold symmetry axis of the encapsulin shell and highlights a potential route for the entry of iron into the encapsulin and towards the active site of EncFtn (**Figure 5.7.3**). In this bacterial compartment model, Fe(II) can enter the encapsulin shell through the surface pore and moves along the well-aligned central channel of EncFtn before accessing the metal ion entry site. is then stripped of it hydration shell and translocated to the FOC where it is and oxidised. The Fe(III) is subsequently released to the outer surface (exit site) of the EncFtn. Fe(III) is then mineralised within the lumen of the encapsulin cage (**Figure 5.7.3 B**). This iron storage model suggests that iron moves from the putative entry site to the exit site of the EncFtn. To validate this hypothesis, the iron oxidation and storage of entry/exit site mutants of Rru-EncFtnSH and Rru-EncFtn-Enc could be detected.

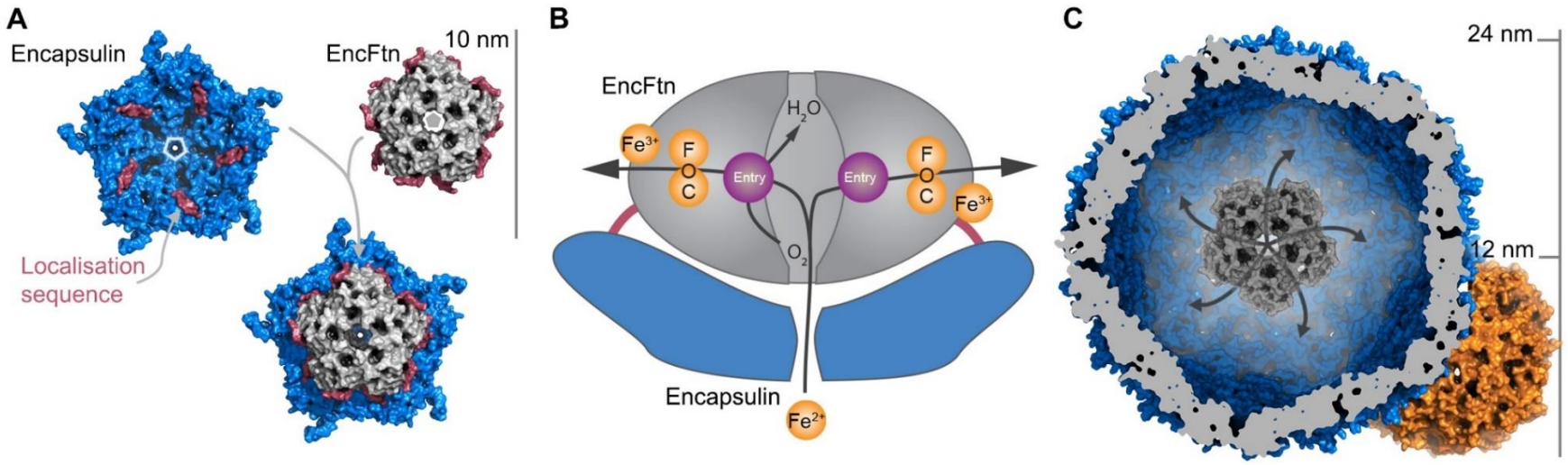


Figure 5.7.3 The Iron storage model of 60-meric encapsulin nanocompartment.

(A) Model of EncFtn (Rru-EncFtnSH, PDB ID: 5DA5) docked into the encapsulin shell. A single pentamer of the icosahedral *T. maritima* encapsulin structure (PDB ID: 3DKT) (Sutter et al. 2008) is shown as a blue surface with the encapsulin localisation sequence of EncFtn shown as a red surface. The C-terminal regions of the EncFtn subunits correspond to the position of the localisation sequences seen in 3DKT. Alignment of EncFtn with 3DKT positions the central channel directly above the pore in the 3DKT pentamer axis (shown as a grey pentagon). (B) Iron metabolic route in encapsulin nanocompartments. As EncFtn is unable to mineralise iron on its surface directly, Fe(II) must pass through the encapsulin shell to access the first metal binding site within the central channel of EncFtn (entry site) prior to oxidation within the FOC and release as Fe(III) to the outer surface of the protein, where it can be mineralised within the lumen of the encapsulin cage. (C) Surface view of EncFtn within the encapsulin nanocompartment (grey and blue respectively). The lumen of the encapsulin nanocompartment is considerably larger than the interior of ferritin (shown in orange behind the encapsulin for reference) and thus allows the storage of significantly more iron. The proposed pathway for iron movement through the encapsulin shell and EncFtn FOC is shown with arrows.

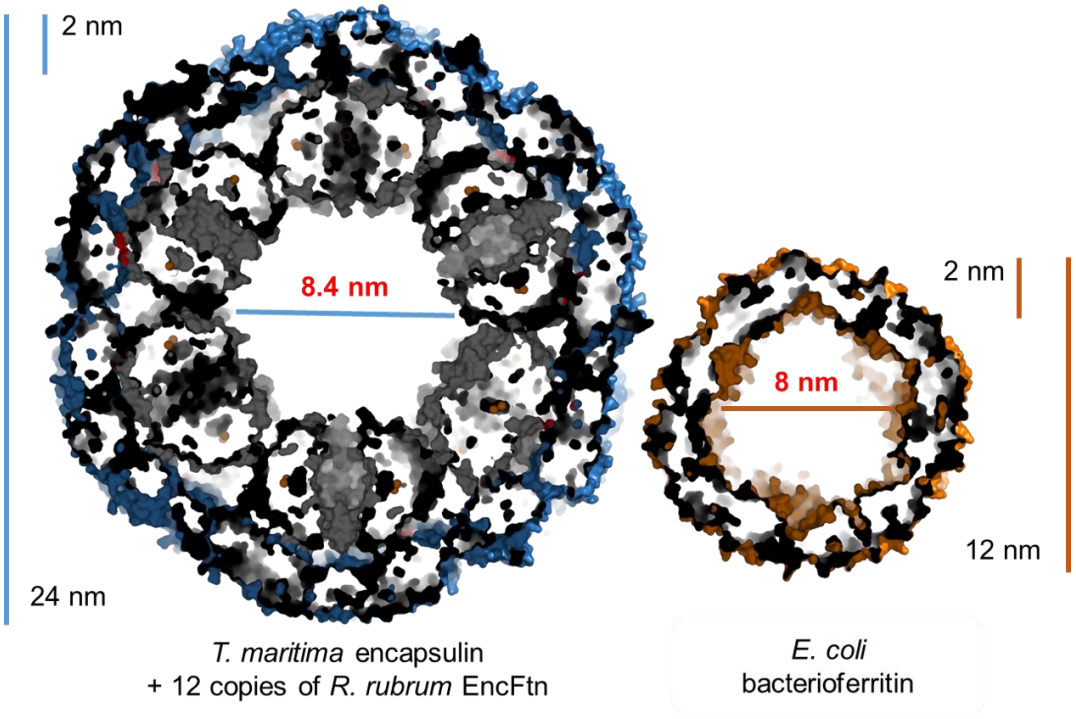


Figure 5.7.4 Interior space comparison of fully occupied encapsulin nanocompartment and *E. coli* bacterioferritin.

Twelve copies of *R. rubrum* EncFtn (Rru-EncFtnSH) coloured in grey surface mode were manually docked into the *T. maritima* encapsulin structure (PDB ID: 3DKT) (Sutter et al. 2008) (blue surface) to produce a model of a fully occupied encapsulin nanocompartment. This arrangement has an interior cavity with around 8.4 nm diameter. *E. coli* bacterioferritin (EcBfr, PDB ID: 1BFR) (Dautant et al. 1998) is a 12 nm octahedron with an 8 nm diameter cavity.

A comparison of the encapsulin nanocompartment and the ferritin nanocage highlights the size differential between the two complexes (**Figure 5.7.3 C**), highlighting the observation that encapsulin nanocompartments can physically allow more iron to be stored within their cavities. However, this model assumes that the 60-meric encapsulin houses only one decamer of EncFtn. When the encapsulin is fully occupied by twelve EncFtn decamers, the inner cavity of the encapsulin nanocompartment is similar to the classical ferritin cage (**Figure 5.7.4**). This manual docking process has been described in **Section 2.11**, and can be validated by an alternative program 'HADDOCK' (Van Zundert et al. 2016). The docking results can also be evaluated using PDBePISA server to analyse the EncFtn-Enc interaction or EncFtn-EncFtn interaction (Krissinel & Henrick 2007). An estimation of the encapsulin nanocompartment interior volume leads to 3,550 nm^3 of 60-meric Tm-Enc with a decameric Rru-EncFtn (**Figure 5.7.3**

C); a fully occupied 60-meric encapsulin nanocompartment possesses about 270 nm^3 interior space, which is analogous to 24-meric classical ferritins (**Figure 5.7.4**).

When *R. rubrum* EncFtn and encapsulin were coexpressed in *E. coli*, it was very likely that the recombinant encapsulin was fully occupied as discussed in **Section 5.5.1**. Therefore, the presence of sixty FOCs in twelve EncFtn decamers means that they could potentially achieve a higher rate of iron mineralisation across the entire nanocompartment than the ferritin cage with 24 FOCs. This has been confirmed by seven-fold difference in iron loading of EncFtn-Enc in our biochemical assays than the horse spleen apoferritin in the same time period (**Figure 5.6.3**). As the iron storage ability depends on iron oxidation rate and the mineralisation cavity, the practical maximum iron loading capacity of encapsulin nanocompartments is still unknown due to the uncertain stoichiometry of EncFtn and encapsulin in native bacteria. When the encapsulin is fully occupied by EncFtn, or the monomer ratio of EncFtn: Enc is 2:1 (as in our model), the iron oxidation rate would reach the maximum, while the iron storage space was limited to minimum (**Figure 5.7.4**). If the encapsulin contains only one decameric EncFtn (EncFtn: Enc = 1:6), the iron oxidation rate is limited by the number of FOCs, while the mineralisation space is maximised (**Figure 5.7.3 C**). It is anticipated that the theoretical maximum iron loading capacity can be achieved by optimising the expression ratio of EncFtn and encapsulin. Whether the ratio of EncFtn and encapsulin is regulated in the native system is not known, but it could be readily achieved in recombinant systems using customised ribosome binding sites, or tunable promoters for future analysis.

5.8 Conclusions and discussion

The recombinant *R. rubrum* EncFtn (Rru-EncFtn) is an active ferroxidase; and its iron oxidation ability was retained when co-expressed with the encapsulin protein Rru-Enc or Rr-Enc. Mutation of the putative ferroxidase centre of Rru-EncFtn reduces the ferroxidase activity of this enzyme, highlighting the essential role of the ferroxidase centre (FOC) in achieving the iron oxidation. Rru-EncFtn is able to bind only 48 iron ions due to its open structure; and while the encapsulin shell is not active to oxidise iron, it is able to bind up to 2,200 Fe atoms. The coexpressed EncFtn and encapsulin (Rru-EncFtn-Enc) is able to bind around 4,150 iron ions in our experimental system, which is significantly more than the calculated combined maximum iron binding (2,776 ± 63) of Rru-EncFtnSH and Rru-Enc as two separate proteins. These results suggest that the Rru-Enc encapsulin protein plays an essential supporting role as the iron storage vessel for the EncFtn protein.

Comparison of the quaternary structure of EncFtn and classical ferritins shows an everted FOC and highlights the importance of a closed-compartment for iron mineralisation. With its open 'ring-doughnut' structure, EncFtn requires localisation within an encapsulin nanocompartment to function as a competent iron store. Collectively, a metabolic model for the storage of iron in encapsulated ferritin nanocompartments is proposed: Fe(II) moves inside encapsulin shell through surface pores to access the EncFtn, followed by translocation from the entry site to the FOC where Fe(II) is oxidised into Fe(III), which is subsequently released into encapsulin cavity and forms iron minerals.

There are a number of open questions that remain to be answered regarding iron storage within encapsulin nanocompartments:

1. The proposed iron storage model should be confirmed through mutagenesis of residues along the proposed oxidation and storage route in the encapsulated ferritin and encapsulin protein.

2. The biochemistry and kinetics of iron oxidation catalysed by Rru-EncFtn remains unclear with respect to the stoichiometry of iron ions, oxidant, and electron transfer. Therefore, the ferroxidase activity should be further explored using multiple

methods including O_2 or peroxide uptake measured by electrode oximetry and stopped-flow kinetic measurement (Bou-Abdallah et al. 2004; Bou-Abdallah et al. 2005; Pfaffen et al. 2013). The transferrin-A_{460} method can be an alternative to the direct measurement of ΔA_{315} regarding ferroxidase activity (Bakkers & Boyerg 1986).

3. What is the iron mineral produced by the encapsulin nanocompartments? Energy Dispersive X-Ray Analysis (EDX) can be applied to identify the composition of iron minerals in encapsulin nanocompartments (Yoshimura 2006). Optimised ICP-MS could be used to determine the stoichiometry of iron to phosphate if it is in the minerals, as has been suggested for the *Myxococcus xanthus* encapsulin nanocompartment (McHugh et al. 2014).

4. Understanding the Rru-Enc structure will give insights into the details of the iron storage model of encapsulin nanocompartments. Although Rru-Enc crystal trials failed to generate high-quality diffraction data, X-ray data collection at room temperature has not yet been attempted to assess the effect of cryoprotection on the crystals. Furthermore, optimised purification methods may yield better diffracting crystals. Besides, single-particle cryo-electron microscopy, as an alternative approach, can be employed to determine the encapsulin Rru-Enc structure and a preliminary dataset has been collected on samples of this protein.

5. The native stoichiometry of Rru-EncFtn and Rru-Enc can be explored by orbitrap mass spectrometry (Snijder et al. 2014). Through promoter optimisation, the influence of the stoichiometry of EncFtn and encapsulin on iron oxidation and storage capacity can be explored.

Chapter 6 Structural characterisation of EncFtn homologues

6.1 Introduction

Encapsulin nanocompartments with encapsulated ferritin-like proteins (EncFtn) are widely distributed in eubacteria and archaea (Giessen & Silver 2016); and there is no clear environmental or lifestyle niche connecting the species with these nanocompartments. The organisation of the encapsulin nanocompartment operon is well conserved with the gene encoding EncFtn upstream of the encapsulin gene. This is the case for the encapsulin nanocompartment operon in eubacteria *R. rubrum*, *T. maritima* and *H. ochraceum* (**Figure 6.1.1**).

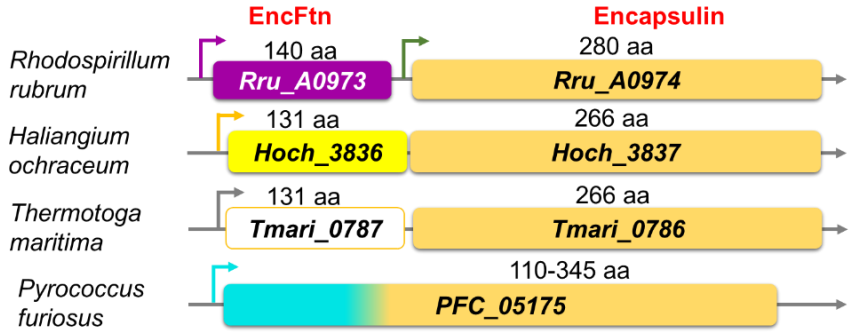


Figure 6.1.1 Genomic context of encapsulin nanocompartments.

Encapsulin nanocompartments including proteins EncFtn and encapsulin are encoded by two neighbouring genes with *EncFtn* locating upstream of *encapsulin* in eubacteria, of which the linker sequence between genes *EncFtn* and *encapsulin* differs in lengths. The arrows indicate the promoter of each gene expression. The EncFtn and encapsulin proteins are encoded by genes *Rru_A0973* and *Rru_A0974* in *R. rubrum*, *Hoch_3836* and *Hoch_3837* in *H. ochraceum*, *Tmari_0787* and *Tmari_0786* in *T. maritima*, respectively. The encapsulin nanocompartments in archaea are represented by *P. furiosus*, in which a single gene *PFC_05175* encodes a polypeptide that contains EncFtn and encapsulin domains.

T. maritima is an anaerobic thermophilic eubacterium and grows optimally at 80°C (Huber et al. 1986). The structure of the encapsulin protein from the encapsulin nanocompartment of *T. maritima* was determined in 2008 as a 60-meric icosahedron with $T = 1$ (PDB ID: 3DKT) (Sutter et al. 2008). *H. ochraceum* is an aerobic, obligate halophile which grows in 0.5–6% NaCl (optimally in 1–3% NaCl) and at 20–40°C (Fudou et al. 2002). The DNA sequence between *EncFtn* and *encapsulin* genes varies in length, which also affects the presence of the promoter of *encapsulin* gene. In *R.*

rubrum a promoter was predicted in the linker sequence (87 bp) between *Rru_A0973* and *Rru_A0974* by using an online program BPROM (Solovyev & Salamov 2011). There are only two bases between *Hoch_3836* and *Hoch_3837* while between *Tmari_0787* and *Tmari_0786* there are 14 bp; in either case no promoter of encapsulin protein was predicted.

In contrast to this conserved genomic architecture the encapsulin nanocompartment in *P. furiosus* is encoded by a single gene *PFC_05175* (**Figure 6.1.1**). *P. furiosus* is an anaerobic hyperthermophilic archaeon and grows optimally at about 100°C and pH 7 (Namba et al. 2005). The crystal structure of the encapsulin nanocompartment isolated from *P. furiosus* displayed a 30–32 nm icosahedron with electron density observed only for the encapsulin shell from residues 110 to 345 of the mature protein (Akita et al. 2007). This 'one gene encoding an encapsulin nanocompartment' has been found in about 10% archaeal species according to the orthologue search of protein PFC_05175 in KEGG database (Ogata et al. 1999).

Given the different environmental niches that these microorganisms live it is interesting to know if they possess different structural and functional features of EncFtn related to their lifestyles. In this chapter, truncated EncFtn from *H. ochraceum, T. maritima* and *P. furiosus* was overexpressed in *E. coli* and purified through IMAC and SEC. The structures of these recombinant proteins were determined using X-ray crystallography. The structural features of EncFtn homologues will be further discussed related to the ecological niches of their microbes. As *T. maritima* EncFtn failed to grow any protein crystals, the protein was not analysed further.

6.2 Protein purification

Rru-EncFtn shares 55% amino acid sequence identity with Hoch-EncFtn (Hoch_3836) and 29% identity with the EncFtn domain of PFC_05175 (Pfc-EncFtn). A multiple sequence alignment of EncFtn homologues reveals the highly conserved ferroxidase centre including E32, E62, H65 and Y39 of Rru-EncFtnSH. While the metal ion entry site (E31 and E34 of Rru-EncFtnSH) and exit site (H57, E61 and E64 of Rru-EncFtnSH) are well conserved in Hoch-EncFtn, but poorly conserved in the Pfc-EncFtn. The alignment also showed that the C-terminal localisation sequence shared

by EncFtn family is missing in PFC_05175 (**Figure 6.2.1**). Based on the alignment, constructs for the heterologous expression of Hoch-EncFtn and Pfc-EncFtn were designed for crystallisation as truncated proteins with 1–98 residues (Hoch-EncFtnS) and 1–99 residues (Pfc-EncFtnS), respectively (**Figure 6.2.1**).

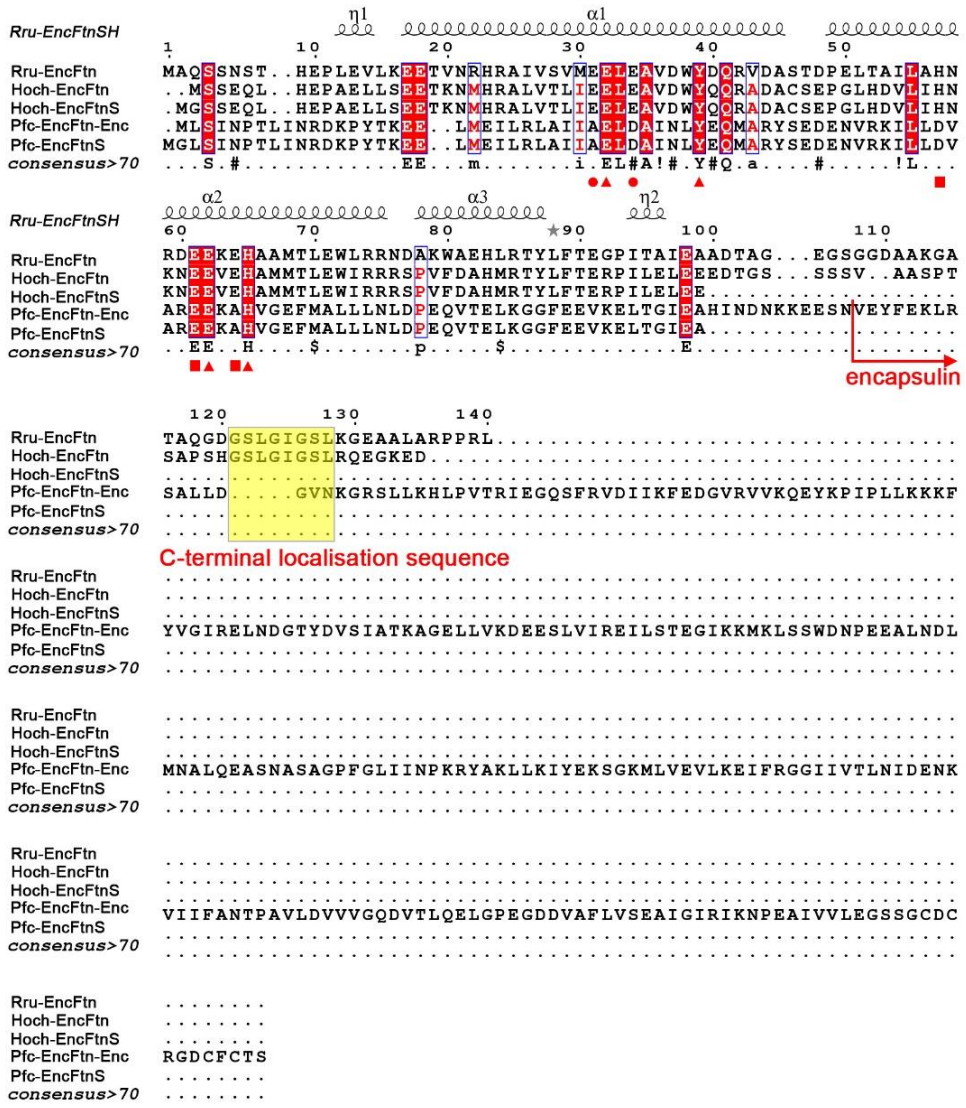


Figure 6.2.1 Protein sequence alignment of EncFtn proteins.

EncFtn proteins in different constructs including Rru-EncFtn (Rru_A0973), Rru-EncFtnSH, Hoch-EncFtn (Hoch_3836), Hoch-EncFtnS, Pfc-EncFtnS and Pfc-EncFtn-Enc (PFC_05175) were aligned using Clustal Omega (Sievers et al. 2011; McWilliam et al. 2013; Li et al. 2015), and the alignment picture was generated by Espript 3.0 (Robert & Gouet 2014). Dots indicate gaps in the alignment. Secondary structure information of Rru-EncFtnSH (PDB ID: 5DA5) and PFC_05175 (PDB ID: 2E0Z) is highlighted on top and in the bottom of the alignment, respectively. The C-terminal localisation sequences of EncFtn are highlighted with yellow rectangles. Residues in the FOC of Rru-EncFtnSH are highlighted with solid red triangles. Residues in the metal ion entry site are pointed by solid red spheres. Residues in the metal ion exit site are underscored by solid red squares.

6.2.1 Pfc-EncFtnS purification

The Pfc-EncFtnS protein construct was calculated to have a molecular weight of 11.304 kDa and pI value of 4.49. It was purified from *E. coli* Tuner(DE3) cells grown in LB medium using Q sepharose AEC (**Figure 6.2.2**), followed by SEC (**Figure 6.2.3**). Unlike the size-exclusion chromatogram of Rru-EncFtnSH, Pfc-EncFtnS only displayed a single A_{280} peak and an A_{315} peak at around 72 ml retention volume, suggesting the presence of a single oligomeric state with iron bound. The SDS-PAGE analysis revealed a major band at decamer size (around 110 kDa), which was later identified as PFC_05175 protein through PMF. It suggested a stronger interaction in the decamer assembly of Pfc-EncFtnS.

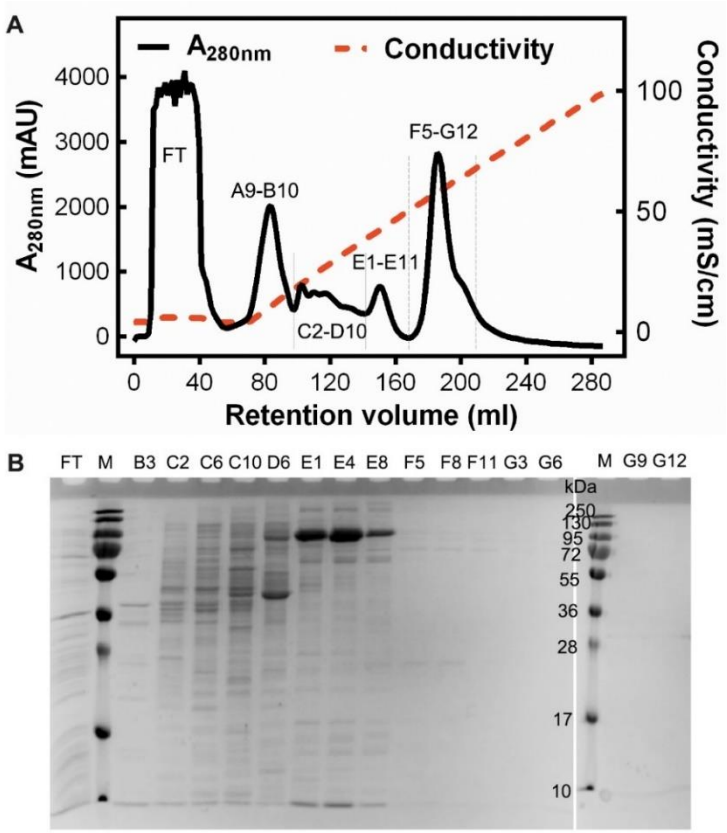


Figure 6.2.2 Anion exchange chromatography purification of Pfc-EncFtnS using a Q sepharose column.

(A) Q-sepharose purification chromatogram of Pfc-EncFtnS. Fractions FT indicate the flow-through. Proteins were eluted by a linear gradient of 0 - 1 M NaCl over 20 column volumes as fractions from A9 to G12. (B) Fractions from (A) were analysed by 15% SDS-PAGE in which only fractions E1-E8 showed around 130 kDa band, corresponding to the decamer size of Pfc-EncFtnS.

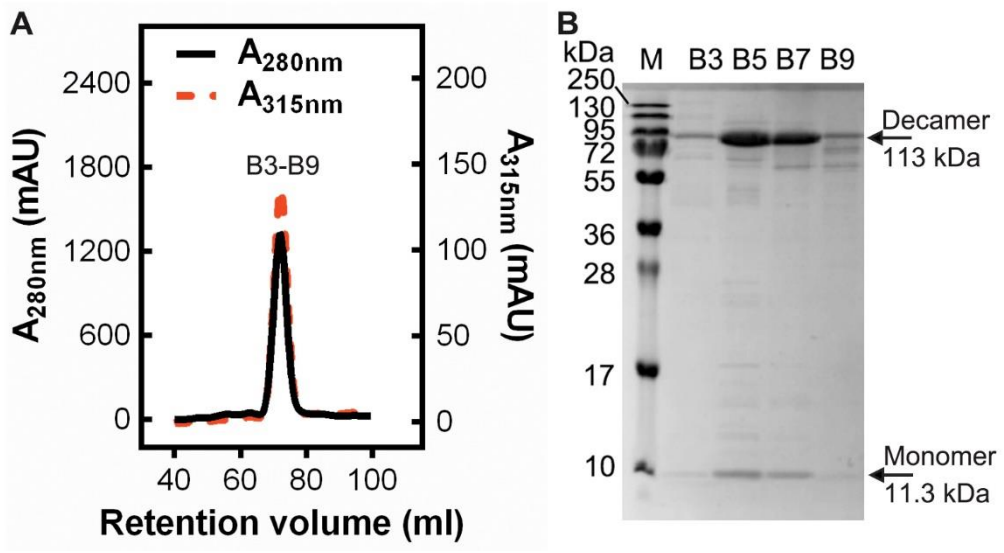


Figure 6.2.3 Size-exclusion chromatography purification of Pfc-EncFtnS using HiLoad Superdex 16/600 200 pg column.

(A) Size exclusion chromatogram of Pfc-EncFtnS purified from *E. coli* Tuner(DE3). Proteins were eluted by GF buffer as fractions B3-B9. (B) Fractions from (A) were analysed by 18% SDS-PAGE as a minor monomer-sized band (11.3 kDa) and a major decamer-sized band (113 kDa).

6.2.2 Hoch-EncFtnS purification

Hoch-EncFtnS protein was calculated to have a molecular weight of 11.572 kDa and pI of 4.76 by ExPASy ProtParam tool (Gasteiger et al. 2005). It was purified from *E. coli* BL21(DE3) grown in LB medium by Q sepharose AEC (**Figure 6.2.4**), followed by SEC (**Figure 6.2.5**). The gel-filtration chromatogram of Hoch-EncFtnS displayed a number of A_{280} and A_{315} peaks at around 47, 60, 69, 78 and 88 ml retention volume. The SDS-PAGE analysis revealed more than two bands of monomer and dimer masses, suggesting the presence of multiple oligomeric states of Hoch-EncFtnS in solution. It was very interesting to observe that the A_{315} curve showed a similar profile with A_{280}, which is in contrast to the observation that monomeric Rru-EncFtnSH has no absorbance at 315 nm due to the lack of iron binding. However, since the A_{315} had the highest peak at 110 ml, of which no protein was detected, the solution might have contributed to the A_{315} of the Hoch-EncFtnS purification.

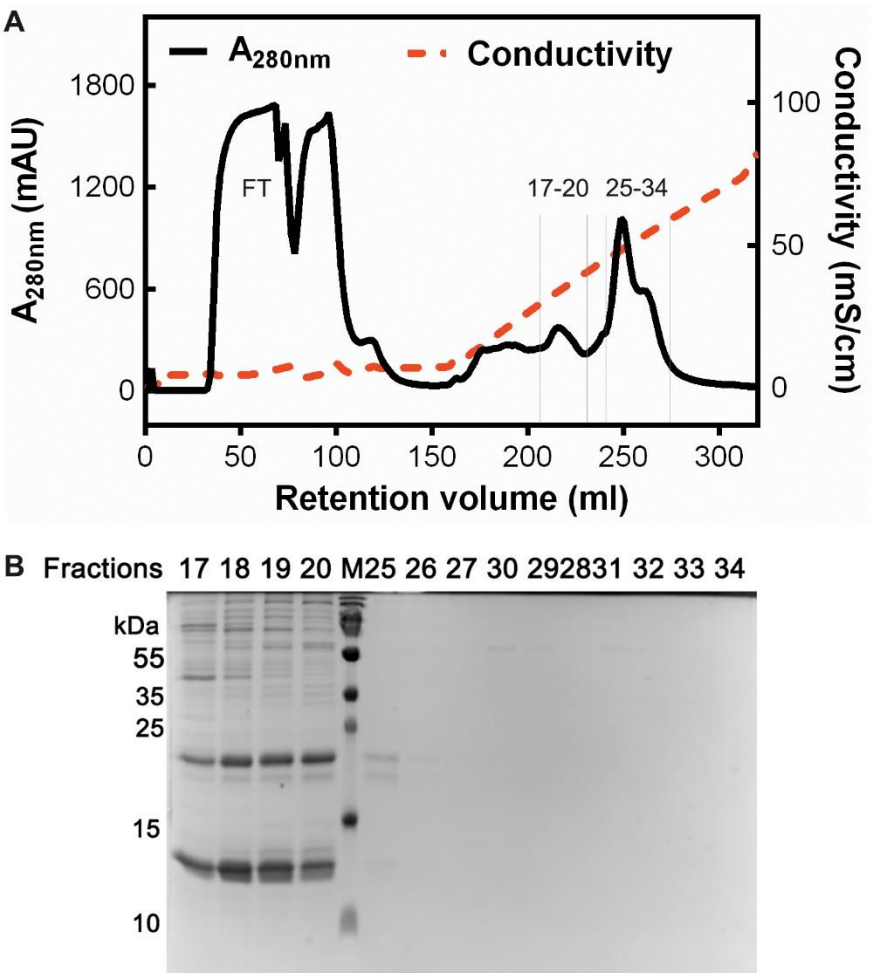


Figure 6.2.4 Anion exchange chromatography purification of Hoch-EncFtnS using Q sepharose column.

(A) Q sepharose purification chromatogram of Hoch-EncFtnS. Fractions FT indicates the flow-through. Proteins were eluted by a NaCl gradient from 0–1 M over 20 column volumes. (B) Fractions 17-34 collected from (A) were analysed by 18% SDS-PAGE in which only fractions 17-20 show two major bands around 12 and 25 kDa, correlating with the estimated monomer and dimer mass of Hoch-EncFtnS, respectively.

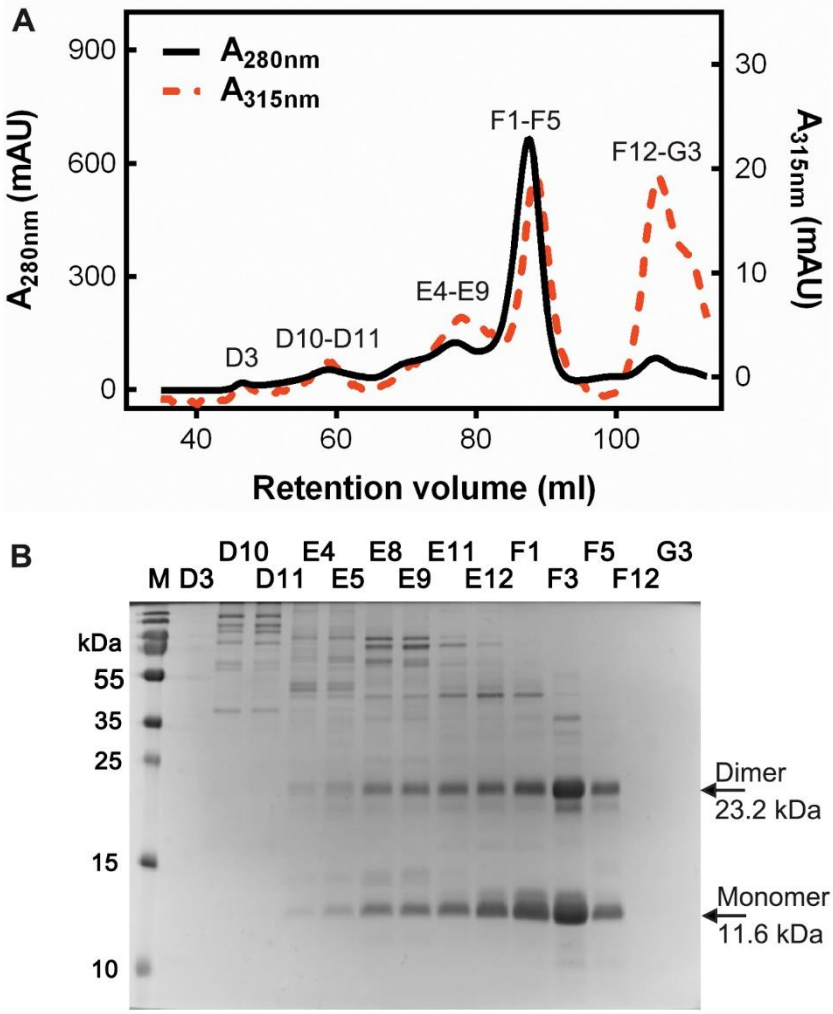


Figure 6.2.5 Size exclusion chromatography purification of Hoch-EncFtnS using HiLoad Superdex 16/600 200 pg column.

(A) Size-exclusion chromatogram of Hoch-EncFtnS purified from *E. coli* BL21(DE3). Proteins were eluted by GF buffer as fractions D3-G3. (B) 15% SDS-PAGE analysis of S200-purified Hoch-EncFtnS displays two major bands approximately monomer and dimer masses.

6.3 Crystallisation and structure determination of Hoch-EncFtnS and Pfc-EncFtnS

The major size-exclusion chromatographic peak fractions of Pfc-EncFtnS and Hoch-EncFtnS were pooled separately and concentrated to around 10 mg/ml before being subjected to crystallisation trials. The Pfc-EncFtnS fractions crystallised easily in many conditions from JCSG plus (Molecular Dimensions) (Newman et al. 2005) and PEG/Ion (Hampton Research) screening kits using sitting-drop vapour diffusion method. The best diffraction data were collected from a crystal grown in 0.2 M lithium sulphate monohydrate and 20% (w/v) PEG 3350 (**Figure 6.3.1**). Cryo-protectant solution including 0.2 M lithium chloride, 20% (w/v) PEG 3350 and 15% (v/v) PEG 200 was used when flash cooling the crystal prior to the data collection at 100 K at the Diamond Light Source (Didcot, UK). Hoch-EncFtnS crystallised in 0.2 M sodium chloride, 0.1 M Bis-Tris (pH 5.5) and 20% (w/v) PEG 3350 using the hanging-drop vapour diffusion method (**Figure 6.3.1**). 15% (v/v) PEG 200 was supplemented into the mother liquor as the cryo-protectant solution.

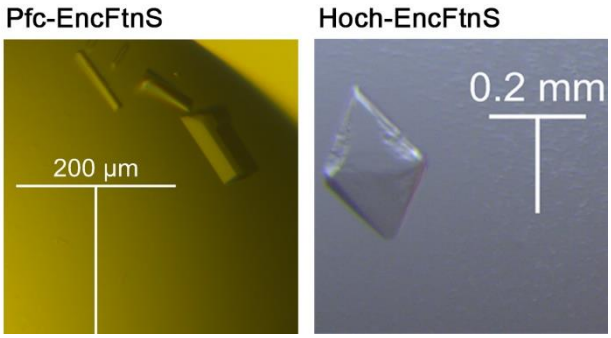


Figure 6.3.1 Crystal images of Pfc-EncFtnS and Hoch-EncFtnS.

The horizontal and vertical scale bars indicate the crystal size.

Both diffraction datasets were collected at 10–100 eV above the theoretical Fe-K_α edge, integrated and scaled using XDS , and merged with Aimless (Evans & Murshudov 2013). Structures were phased by molecular replacement using a modified Rru-EncFtnSH monomer or decamer structure as the search model in Phaser (McCoy et al. 2007). The structures were then built using Coot (Emsley et al. 2010) and refined with Refmac5 (Pannu et al. 1998; Steiner et al. 2003), followed by Phenix.refine (Adams et al. 2010). The data collection and refinement statistics are shown in **Table 6.3.1**.

	Hoch-EncFtnS	Pfc-EncFtnS
Data collection		
Beamline	I04	I04
Wavelength (Å)	1.72	1.74
Resolution range	47.82–2.06 (2.13–2.06)	47.19–2.03 (2.10–2.03)
Space group	*C* 1 2 1	*P* 1 2_1 1
Unit cell (Å) a	91.63	99.85
b	92.65	110.06
c	119.29	136.27
α = γ = 90°, ß =	106.93	91.22
Total reflections	112,239 (10,771)	371,374 (36,434)
Unique reflections	57,293 (5,570)	190,563 (18,927)
Multiplicity	2.0 (1.9)	1.9 (1.9)
Completeness (%)	96 (94)	100 (99)
Mean I/sigma(I)	9.50 (2.48)	13.87 (5.76)
Wilson B-factor	31.67	22.84
R_{merge}	0.0403 (0.283)	0.0396 (0.158)
R_{meas}	0.0571 (0.401)	0.0560 (0.224)
CC1/2	0.997 (0.914)	0.996 (0.95)
CC*	0.999 (0.977)	0.999 (0.987)
Image DOI		
Refinement		
R_{work}	0.199 (0.285)	0.174 (0.223)
R_{free}	0.244 (0.334)	0.202 (0.245)
CC(work)	0.965 (0.930)	0.967 (0.931)
CC(free)	0.936 (0.855)	0.953 (0.914)
Number of non-hydrogen atoms	7,819	25,735
macromolecules	7,650	23,691
ligands	1	30
Protein residues	916	2,922
RMS(bonds) (Å)	0.003	0.002
RMS(angles) (°)	0.46	0.44
Ramachandran favoured (%)	99	100
Ramachandran allowed (%)	1.2	0.1
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0.84	0.35
Clashscore	1.2	1.21
Average B-factor ($Å^2$)	44.57	28.97
macromolecules	44.59	28.32
ligands	34.68	19.73
solvent	43.7	36.74
Protein Data Bank code		

Table 6.3.1 Data collection and refinement statistics of Hoch-EncFtnS and Pfc-EncFtnS.

Statistics for the highest-resolution shell are shown in parentheses. Friedel mates were averaged when calculating reflection numbers and statistics.

6.4 Structural analysis of Hoch-EncFtnS and Pfc-EncFtnS

The Hoch-EncFtnS (Hoch) structure was determined to 2.05 Å resolution in space group *C* 1 2 1; with ten monomers in an annular decamer in the ASU. The structure of Pfc-EncFtnS (Pfc) was determined to 2.02 Å resolution in space group *P* 1 2_1 1, with three decamers in the ASU. The $RMSD_{C\alpha}$ of Rru-EncFtnSH (Rru) in 16–76 residues (α1-α2) are 0.333 Å with counterparts of Hoch-EncFtnS and 0.461 Å of Pfc-EncFtnS (**Figure 6.4.1** and **Figure 6.4.2**). A closer inspection of the structural alignment revealed three main differences among Hoch, Pfc and Rru structures: the N-terminus, the C-terminal helix and metal binding sites.

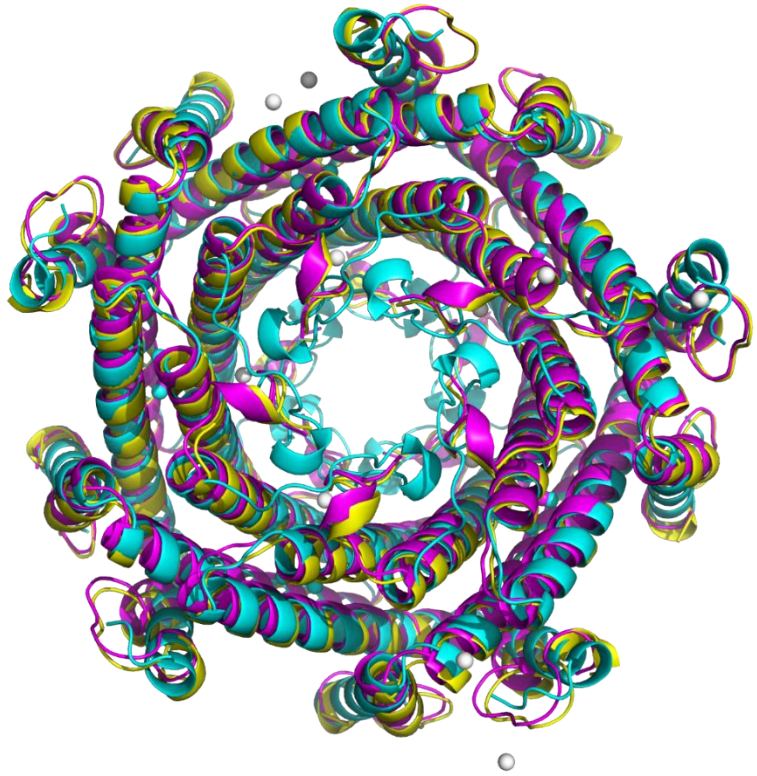


Figure 6.4.1 Decameric structure alignment of EncFtn homologues.

Decamer structures of Rru-EncFtnSH (purple), Pfc-EncFtnS (cyan) and Hoch-EncFtnS (yellow) were aligned by secondary structure matching and are shown in cartoon mode with coordinated iron ions (purple and cyan balls) and calcium ions (white balls).

The N- and C-termini of the Hoch-EncFtnS structure displayed high similarity with Rru-EncFtnSH structure. The crystal structure of the Pfc-EncFtnS construct had an ordered N-terminus, which is in contrast to the disordered N-termini of the Rru-EncFtnSH and Hoch-EncFtnS structures (**Figure 6.4.2**). This observation implied that the N-terminal region of the EncFtn homologues had a varying degree of flexibility in the different species. Since the N-terminus of the protein was found within the central channel of decameric EncFtn that connected the 5-fold symmetry pore of encapsulin and the metal ion entry site of Rru-EncFtn as a potential Fe(II) moving route in encapsulin nanocompartments (**Figure 5.7.3**), the flexibility of the N-terminus might affect the Fe(II) access to the metal ion entry site and thus the catalytic activity of the EncFtn. The C-terminus of Pfc-EncFtnS was longer and more structured than the equivalent parts of Rru-EncFtnSH and Hoch-EncFtnS. The contribution of the C-terminus to the structural stability will be further explored in the **Figure 6.5.3**.

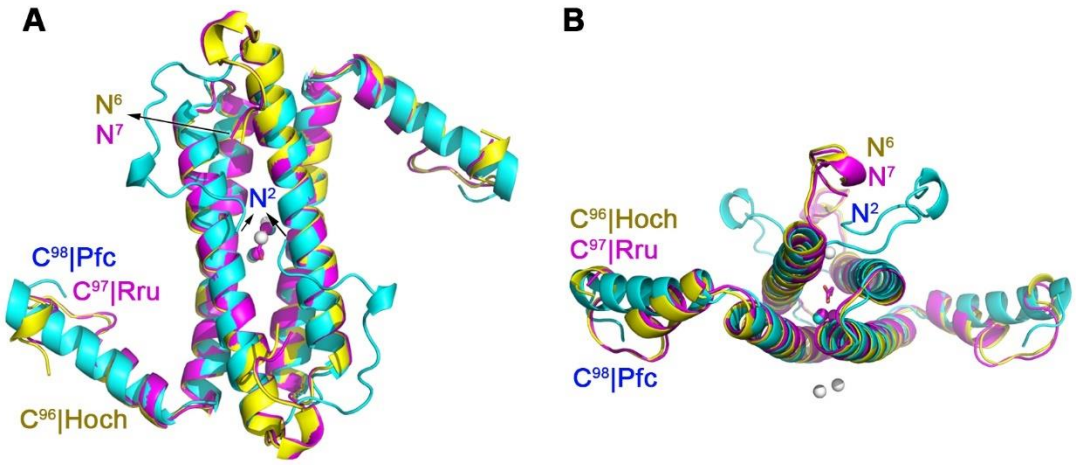


Figure 6.4.2 Structural alignment of FOC dimers of EncFtn homologues.

The aligned FOC dimer structures of Rru-EncFtnSH (Rru, purple), Pfc-EncFtnS (Pfc, cyan) and Hoch-EncFtnS (Hoch, yellow) are shown in cartoon mode with glycolic acid (GOA), iron ions (purple and cyan balls) and calcium ions (white balls) coordinated. GOA, Ca(II) and Fe ions in purple belong to Rru-EncFtnSH structure while iron ions in cyan spheres belong to the Pfc-EncFtnS structure. The N-termini and C-termini of proteins are highlighted in corresponding colours. Panel **B** is the bottom view of panel **A**.

6.4.1 Putative metal binding sites in the Hoch-EncFtnS structure

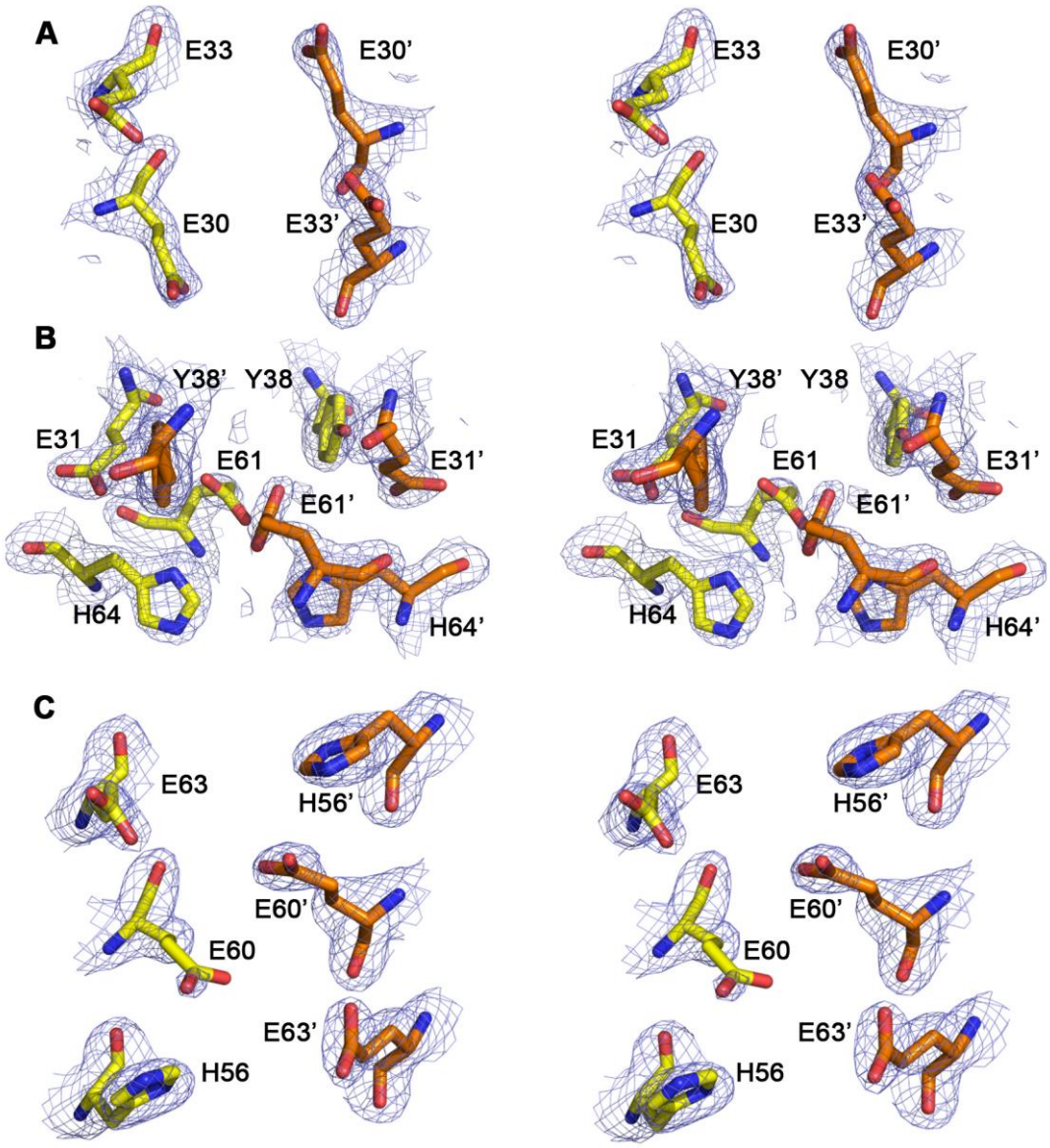


Figure 6.4.3 Wall-eyed stereo view of the equivalent metal binding residues of Rru-EncFtnSH in Hoch-EncFtnS dimer interface.

Protein residues are shown as sticks with yellow and orange carbons for the different chains of Hoch-EncFtnS, green and blue carbons for the different chains of Pfc-EncFtnS structure. 2mFo-DFc density map is superimposed on the structure and contoured at 1.3 σ. (A) The equivalent metal ion entry site of Rru-EncFtnSH in the Hoch-EncFtnS structure consists of E30 and E33 from two chains. (B) The equivalent FOC of Hoch-EncFtnS consists of E31, E61, H64 and Y38 of two adjacent chains. (C) The equivalent metal ion exit site in Hoch-EncFtnS is composed of H56, E60 and E63 of two chains.

Although the metal binding residues of Rru-EncFtnSH are conserved in the Hoch-EncFtnS protein sequence (**Figure 6.2.1**), no metal ions were observed in the equivalent FOC or metal entry/exit sites of Hoch-EncFtnS structure (**Figure 6.4.3**). The metal ion entry site in Hoch-EncFtnS consists of two glutamic acid residues E30

and E33, equivalent to the E31 and E34 of Rru-EncFtnSH structure (**Figure 6.4.7 A**). There was no metal binding in the metal ion entry site of Hoch-EncFtnS structure. The metal ion exit site consisted of H56, E60 and E63 of two chains, equivalent to the H57, E61 and E64 of Rru_A073sH (**Figure 6.4.7 B**). However, the metal ion exit site of Hoch-EncFtnS did not bind any metal in the structure.

The absence of metal in the FOC of the Hoch-EncFtnS structure could be ascribed to two possibilities. The first hypothesis was that Hoch-EncFtnS was not a real ferroxidase and thus could not bind iron. In the FOC of Hoch-EncFtnS structure, two of three main di-iron coordinating residues moved beyond the iron binding distance (Harding et al. 2010). The chi3 angle of E31 in the Hoch-EncFtnS structure is flipped 100° away from the aligned counterparts, and keeps a 4 Å distance with the iron ions of Pfc-EncFtnS and Rru-EncFtnSH structures (**Figure 6.4.7 A**). Besides, the E61 side chain of Hoch moved closer to the iron ions of Pfc-EncFtnS and Rru-EncFtnSH as 1.6 Å. The coordination symmetry also broke in E61 and E61' from Hoch-EncFtnS FOC dimer (**Figure 6.4.6 B**). This hypothesis would be readily validated by ferroxidase assays.

The second hypothesis was that due to the low pH of the Hoch-EncFtnS crystallisation buffer (Bis-Tris, pH 5.5), in which the side chain nitrogen of H64 would possibly be protonated, leading to an inability to coordinate iron and the possibility of hydrogen bonding (3.5 Å) between H64 and E61', thus precluding metal binding completely. To further assess if basic pH of crystallisation conditions combined with iron soaking of crystals will facilitate the iron binding in Hoch-EncFtnS structure, the protein was crystallised in 0.2 M $(NH_4)_2SO_4$, 0.1 M Tris-HCl, pH 8.0, 23% (w/v) PEG 3350 and then soaked in a cryo-protectant solution containing the mother liquor and 12% (v/v) PEG 200 as well as 1 mM acidic $Fe(NH_4)_2(SO_4)_2$ for 1–30 minutes; or crystals grown in similar conditions with 0.1 M Bis-Tris (pH 5.5) instead of Tris-HCl were soaked in the 1 mM acidic $Fe(NH_4)_2(SO_4)_2$ for 1-45 minutes. To reduce the radiation damage caused by free iron ions in crystals, some iron-soaked crystals were also back-soaked swiftly in cryo-protectant without iron. However, none of these crystals displayed di-iron in FOC of Hoch-EncFtnS structures.

6.4.2 Putative metal binding sites in Pfc-EncFtnS

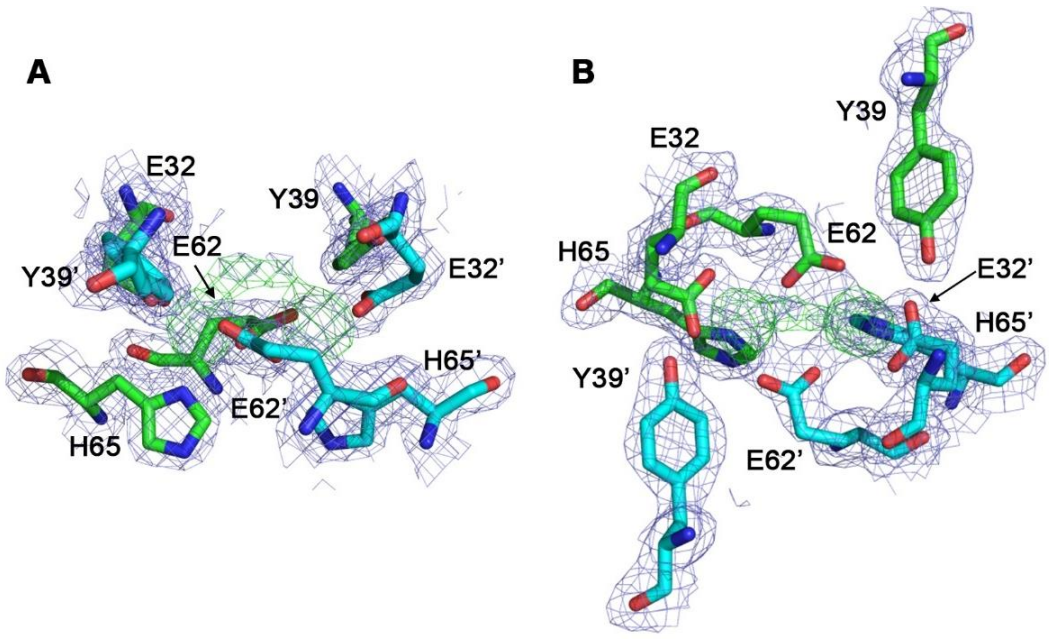


Figure 6.4.4 Putative metal binding sites in Pfc-EncFtnS FOC dimer interface.

(A) Putative metal/ligand binding sites of Pfc-EncFtnS consist of E32, E62, H65 and Y39 of two adjacent chains, which are shown as sticks and coloured in green and cyan. The 2mFo-DFc electron density is shown in blue mesh and contoured at 1.3 σ. The mFo-DFc difference map is shown in green mesh and contoured at 3.5 σ. (B) The top view of (A). Two peaks of the difference map, with each surrounded by monodentate E62, Y39 and bidentate E32', H65' of the adjacent chain, correspond to the two metal ions. There are another two small peaks on top of di-metal, corresponding to water molecules.

The electron density map of the initial model of Pfc-EncFtnS displayed two significant positive peaks in the mFo-DF_C map at the centre of FOC with a difference peak height around 3 σ (**Figure 6.4.4**). The diffraction data were collected at the experimentally determined Fe-*K* edge, and an anomalous difference Fourier map was calculated with the data (**Figure 6.4.5 B**). Despite the Pfc-EncFtnS being produced with no additional iron and the absence of additional iron in the crystallisation mother liquor, the FOC was still occupied by two iron ions. Furthermore, compared to the glycolic acid (GOA) coordinated above the di-iron in Rru-EncFtnSH, two water molecules were coordinated above the di-iron centre in Pfc-EncFtnS (**Figure 6.4.6**). The metal ion entry and exit sites in the Pfc-EncFtnS structure are not well conserved with the Rru-EncFtn and Hoch-EncFtn proteins and did not display any metal binding in the crystal structure (**Figure 6.4.5 A/C** and **Figure 6.4.7**). The entry site in Pfc-EncFtnS contained A31 and D34 in contrast to the two glutamic acid residues found in the other

two homologues. In the absence of suitable coordinating residues, it is perhaps not surprising that no metal is seen in this site in the Pfc-EncFtnS structure. The lack of the conserved secondary metal-binding sites in the Pfc-EncFtnS structure suggests its biochemistry may have different kinetics, or even activity compared to the Rru-EncFtnSH, which is worth exploring in future studies.

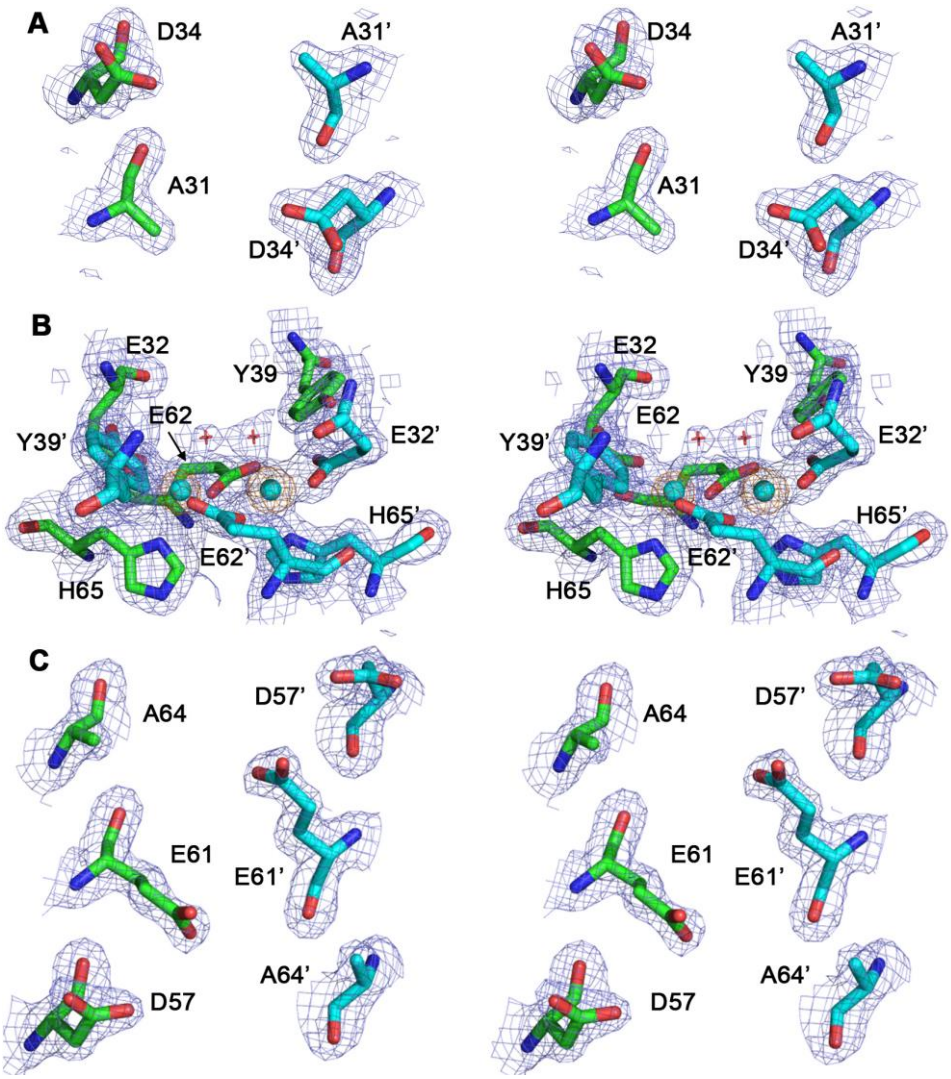


Figure 6.4.5 Wall-eyed stereo view of the equivalent metal binding residues of Rru_0973sH in Pfc-EncFtnS dimer interface.

Protein residues are shown as sticks with green and blue carbons for the different chains of the Pfc-EncFtnS structure. 2mFo-DFc density map is superimposed on the structure and contoured at 1.3 σ. (A) The equivalent metal ion entry site in Pfc-EncFtnS structure consists of A31 and D34 from two chains. (B) The equivalent FOC of Pfc-EncFtnS consists of E32, E62, H65 and Y39 of two adjacent chains, which coordinate di-iron ions (cyan spheres) with two water molecules (red crosses) bound above. The anomalous difference Fourier map obtained from data collected at the Fe-*K* edge is superimposed as an orange mesh and contoured at 5 σ. (C) The equivalent metal ion exit site in Pfc-EncFtnS is composed of D57, E61 and A64 of two chains.

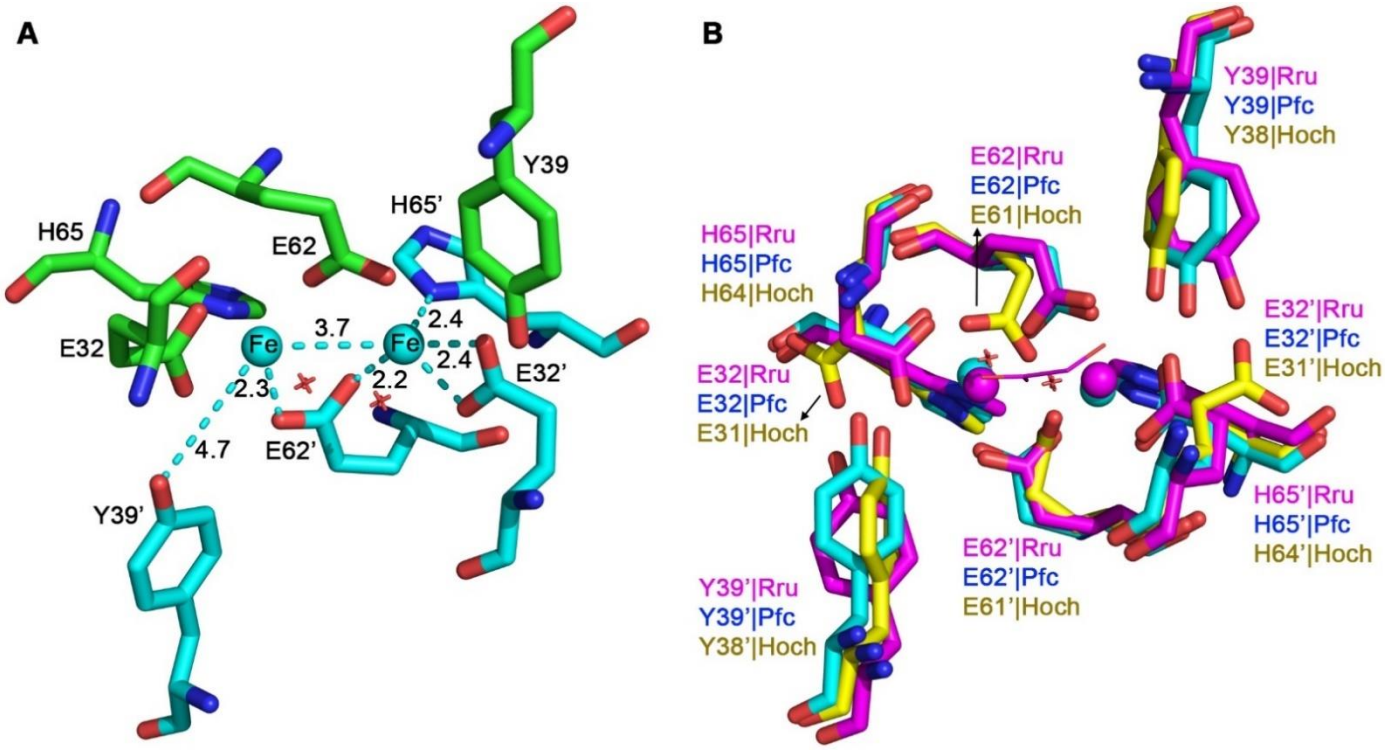


Figure 6.4.6 FOC of Pfc-EncFtnS and the superimposed FOC of EncFtn homologues.

(A) Symmetrical FOC of Pfc-EncFtnS with distance between iron ions and coordinating residues shown in Ångstroms. Protein residues are shown as sticks with green and blue carbons for the different chains of Pfc-EncFtnS structure in the dimer. (B) Aligned FOC of Rru-EncFtnSH (purple), Hoch-EncFtnS (yellow) and Pfc-EncFtnS (cyan) are shown in stick mode. The positions of di-iron ions of Rru-EncFtnSH and Pfc-EncFtnS are well aligned and shown as purple and cyan spheres, respectively. The two water molecules above the cyan di-iron ions of Pfc-EncFtnS are shown in red crosses. Glycolic acid bound by the purple di-iron ions of Rru-EncFtnSH are shown in purple lines. Labels of residues from each protein were in the same colour following by protein anonyms (Rru: Rru-EncFtnSH, Pfc: Pfc-EncFtnS, Hoch: Hoch-EncFtnS). Labels of residues from dimers were differentiated with a prime symbol. Among aligned residues, Hoch differs from Pfc and Rru in E31 and E61 residues, with the chi3 angle of E31 residue flipping away from the centre and E61 side chain moving closer to di-iron of Pfc-EncFtnS as 1.7 Å. Besides, the coordinating symmetry breaks in E61 and E61' of Hoch-EncFtnS.

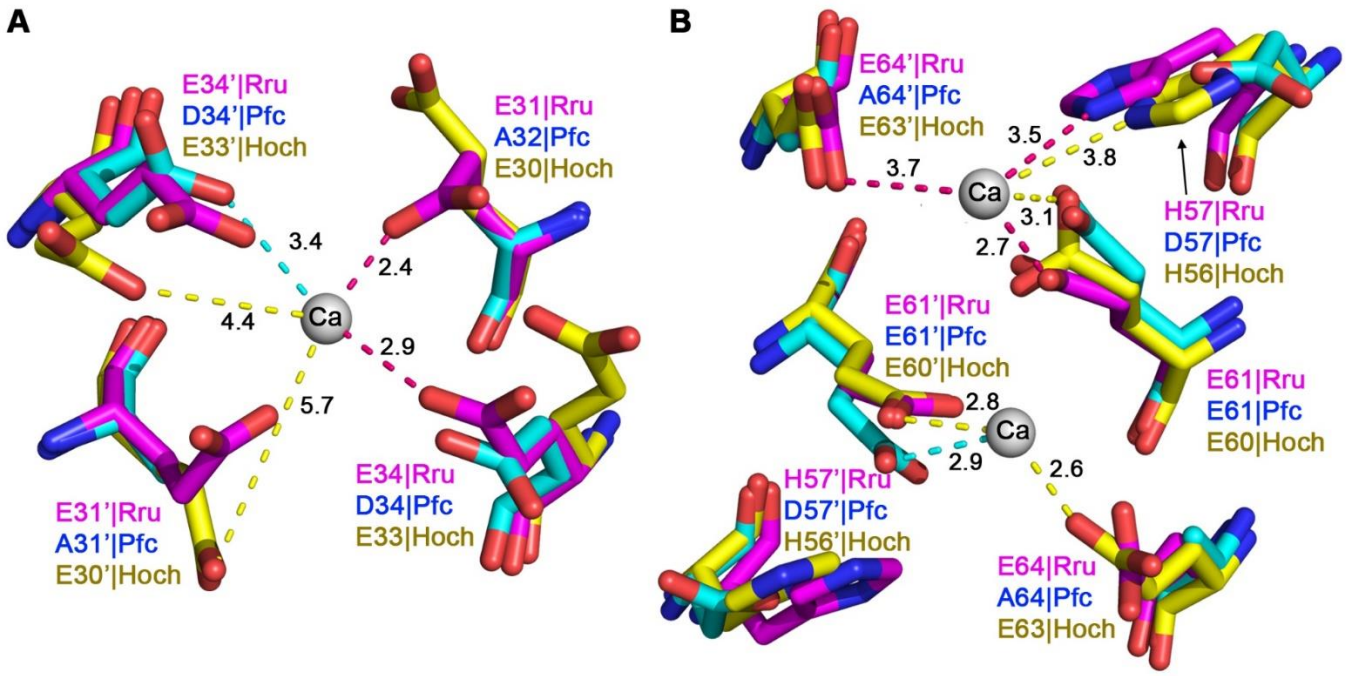


Figure 6.4.7 Superimposed metal ion entry and exit sites of EncFtn homologues.

Aligned metal ion entry sites (A) and exit sites (B) of Rru-EncFtnSH (Rru, purple), Pfc-EncFtnS (Pfc, cyan) and Hoch-EncFtnS (Hoch, yellow) are shown in sticks. Ca(II) ions shown in white spheres are coordinated by Rru-EncFtnSH. Distance between Ca(II) and neighbouring residues labelled by dashed lines of different colours with Å as the unit. The geometry of entry sites in Hoch-EncFtnS and Pfc-EncFtnS suggests that they do not bind metal ions. The exit site coordination implies that Hoch may resemble Rru while Pfc does not bind metal ions in this region.

6.4.3 Surface characteristics of EncFtn homologue structures

To analyse the surface characteristics of the EncFtn homologues, molecular surfaces with mapped electrostatic potentials were calculated using the APBS plugin (Baker et al. 2001) in PyMOL (Schrodinger 2015). The three EncFtn homologues display distinct charge distributions on the decamer surface (**Figure 6.4.8 A–F**). It is very clear to see that the size of five-fold symmetry channel varies among Rru, Pfc and Hoch, although this is dependent on the completeness of the crystallographic model in the N-terminal regions as mentioned above. The five-fold symmetry channel of EncFtn in the native encapsulin *in vivo* is proposed to be the route of entry of Fe(II) from the five-fold channel of the encapsulin towards the metal ion entry site of EncFtn (**Figure 5.7.3**). Notably, the large size of the five-fold symmetry channel seen in Rru-EncFtnSH and Hoch-EncFtnS structures was likely an artefact of the crystallographic models and did not imply greater level of iron access through the complex (**Figure 6.4.8 A/C**), as it was surrounded by the N-terminus of the protein in the EncFtn homologue structures.

The negatively-charged patch in the metal ion entry site of Rru-EncFtnSH was consistent with the structural observation of E31 and E34 residues (**Figure 6.4.8 G**). The metal entry site of Pfc-EncFtnS was blocked by the N-terminus of the protein, which was shown as a positively-charged patch contributed by the backbone nitrogens of residues in this region (**Figure 6.4.8 H**). The red patch of negative charge in the metal ion entry site of Hoch (**Figure 6.4.8 I**) supported the wider space measured in **Figure 6.4.7 A**. The surface comparison of metal ion exit sites of Rru-EncFtnSH and Hoch-EncFtnS revealed a similar negatively-charged region (**Figure 6.4.8 J/L**), while the narrower red patch in Pfc-EncFtnS metal exit site (**Figure 6.4.8 K**) was consistent with the sole conserved E61 observed in the superimposed structures (**Figure 6.4.7 B**).

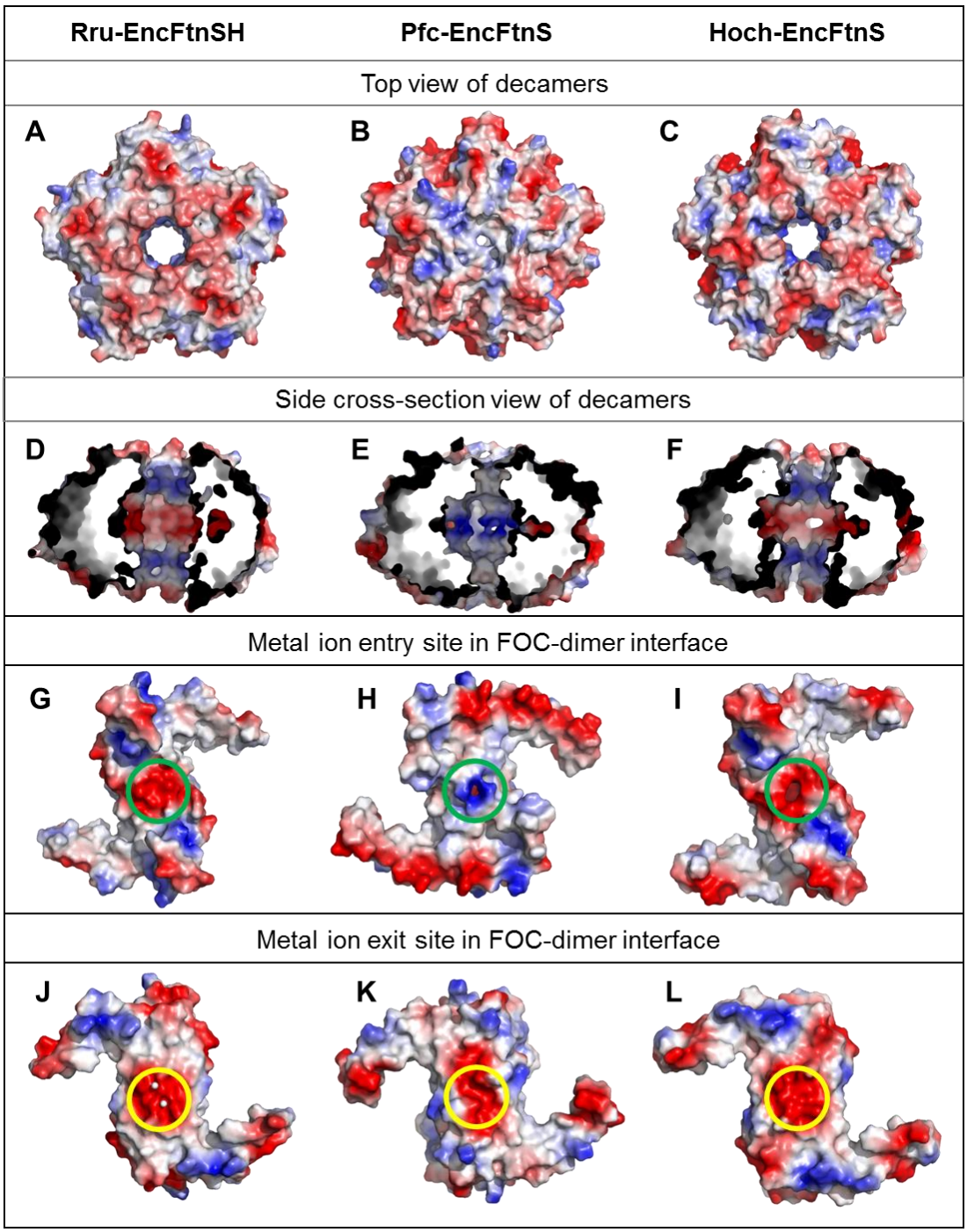


Figure 6.4.8 Electrostatic surfaces of EncFtn homologues.

The solvent accessible surface of EncFtn homologues are shown, coloured by electrostatic potential as calculated in PyMOL(Schrodinger 2015). Negatively charged regions are coloured red and positively charged regions in blue, neutrally charged regions in grey. (A, B, C) Top view of decameric Rru-EncFtnSH, Pfc-EncFtnS and Hoch-EncFtnS. (D, E, F) Side cross-section view of (A, B, C). (G, H, I) Side view of monomeric Rru-EncFtnSH, Pfc-EncFtnS and Hoch-EncFtnS with metal ion entry site highlighted by green circles. (J, K, L) Side view of monomeric Rru-EncFtnSH, Pfc-EncFtnS and Hoch-EncFtnS with metal ion exit site highlighted by yellow circles. Ca(II) as white spheres bound by metal exit site of Rru-EncFtnSH was shown in (G) as well.

6.5 The dimer interface comparison of EncFtn homologues

With a conserved quaternary architecture, both Pfc-EncFtnS and Hoch-EncFtnS have dimer interfaces that correspond to the those seen in the Rru-EncFtnSH structure, with the main difference seen in the metal coordination within the FOC interface (**Figure 6.5.1**). There are no disulphide bridges present in either of these interfaces.

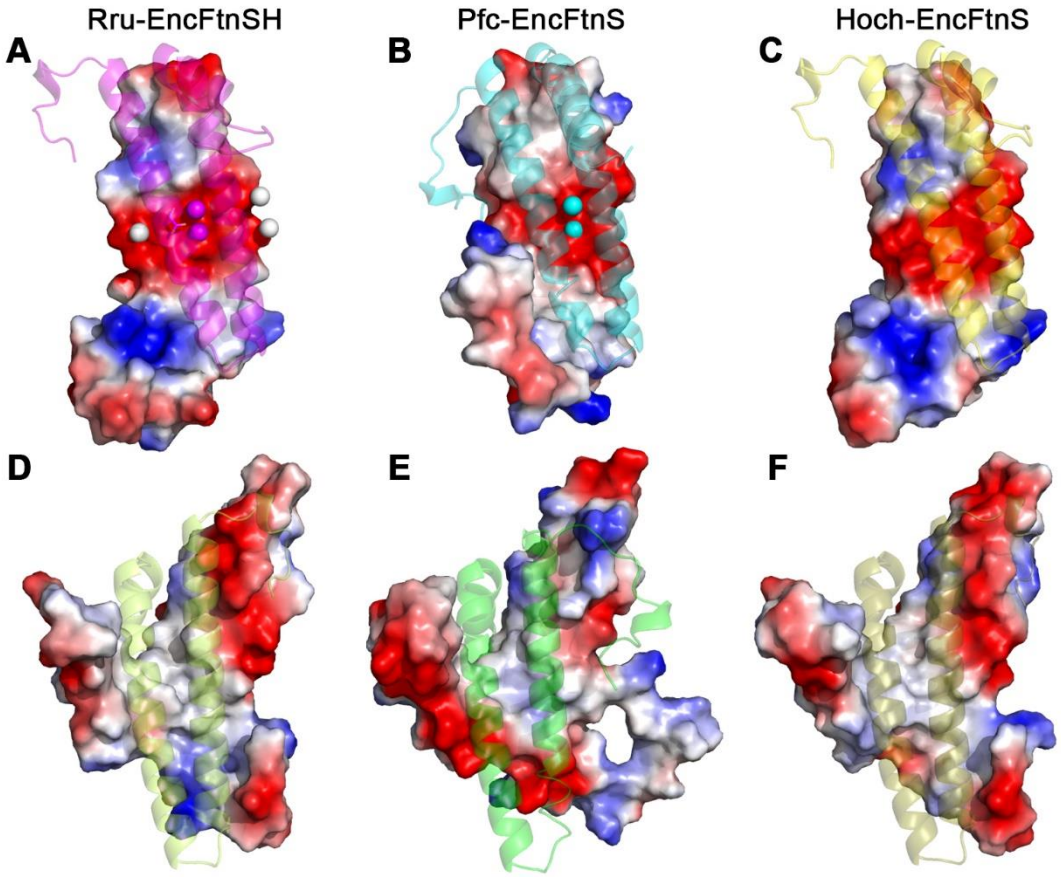


Figure 6.5.1 Dimer interfaces of EncFtn homologues.

The two distinct dimer interfaces of the EncFtn homologues are shown with one chain shown as a secondary structure cartoon and the other chain as a surface representation with mapped electrostatic potential calculated in PyMOL (Schrodinger 2015). On the surface red areas indicate negatively-charge, blue positive charge, and white showing neutral regions. (A) The FOC dimer interface of Rru-EncFtnSH is shown with iron ions bound as magenta spheres and calcium ions as white spheres. (B) The FOC dimer interface of Pfc-EncFtnS is shown with di-iron ions shown in cyan spheres. (C) The FOC dimer interface of Hoch-EncFtnS is shown. Non-FOC dimer interfaces of (D) Rru-EncFtnSH, (E) Pfc-EncFtnS and (F) Hoch-EncFtnS are shown in the bottom row.

6.5.1 Structural adaptation of Pfc-EncFtnS to hyperthermophilic conditions

Pfc-EncFtnS buried 45% surface area in the FOC dimer interface, which was stabilised by 14 hydrogen bonds and 5 salt bridges (**Table 6.5.1** and Figure **6.5.2**). Here the contribution of iron binding to the interface interaction was not counted by PDBePISA server (Krissinel & Henrick 2007). The non-FOC dimer interface of Pfc-EncFtnS comprised around 47% of the buried surface area, which was stabilised by 19 hydrogen bonds and 24 salt bridges (**Figure 6.5.3 B/E** and **Table 6.5.1**). Notably, the more structured C-terminal α3 helix and longer N-terminus of Pfc-EncFtnS allowed it to interact not only with the adjacent chain, but also with the chain next to the neighbouring chain (**Figure 6.5.3 B/E**), which is likely to enhance the structural stability of the decamer. Most thermophilic proteins are generally characterised by large hydrophobic core, increased disulphide bonds, high number of salt-bridges and surface charges compared to the mesophilic counterparts (Reed et al. 2013). In the case of hyperthermophilic Pfc-EncFtnS, more interface residues and hydrogen bonding involved in the FOC dimer interface as well as the higher number of salt bridges in the non-FOC dimer interface than the mesophilic Rru-EncFtnSH counterpart, supporting the notion that thermophiles maintain the structural stability under high temperature via structural adaptation such as increased ionic interactions (Reed et al. 2013).

Proteins	FOC dimer interface			Non-FOC dimer interface		
	Interface residues	Number of hydrogen bonds	Number of salt bridges	Interface residues	Number of hydrogen bonds	Number of salt bridges
Rru-EncFtnSH	34%–36%	8	6	60%	21	15
Pfc-EncFtnS	45%	14	5	47%	19	24
Hoch-EncFtnS	34%	5	15	58%	17	4

Table 6.5.1 Dimer interface comparison of EncFtn homologues analysed by PDBePISA.

The FOC dimer and non-FOC dimer of Rru-EncFtnSH, Pfc-EncFtnS and Hoch-EncFtnS were analysed by PDBePISA to determine the residues and bonds formed in the interfaces (Krissinel & Henrick 2007).

6.5.2 Structural adaptation of Hoch-EncFtnS to halophilic conditions

Hoch-EncFtnS resembled Rru-EncFtnSH regarding the residues forming the dimer interfaces (**Table 6.5.1**). The FOC dimer interface contained 5 hydrogen bonds and 15 salt bridges (Figure **6.5.2 C/F**). The non-FOC dimer interface was stabilised by 17 hydrogen bonds and 4 salt bridges (**Figure 6.5.3 C/F**). Halophilic proteins that are able to stay active in hypersaline environment are characterised by increased negative surface charge due to increased acidic residues (Siglioccolo et al. 2011), reduced numbers of hydrophobic residues as well as the salt-dependent assembly (Reed et al. 2013). In the case of halophilic Hoch-EncFtn, both the full length protein and the short construct contained the same amount of acidic amino acid residues with the counterparts of Rru-EncFtn. The Hoch-EncFtn full length protein consisted of 59 non-polar residues by contrast to the 70 of Rru-EncFtn according to the estimation by EMBOSS Pepstats (Rice et al. 2000). Since EncFtn is natively enclosed in its encapsulin protein, the structural adaptation of encapsulated proteins may be different from the free enzymes.

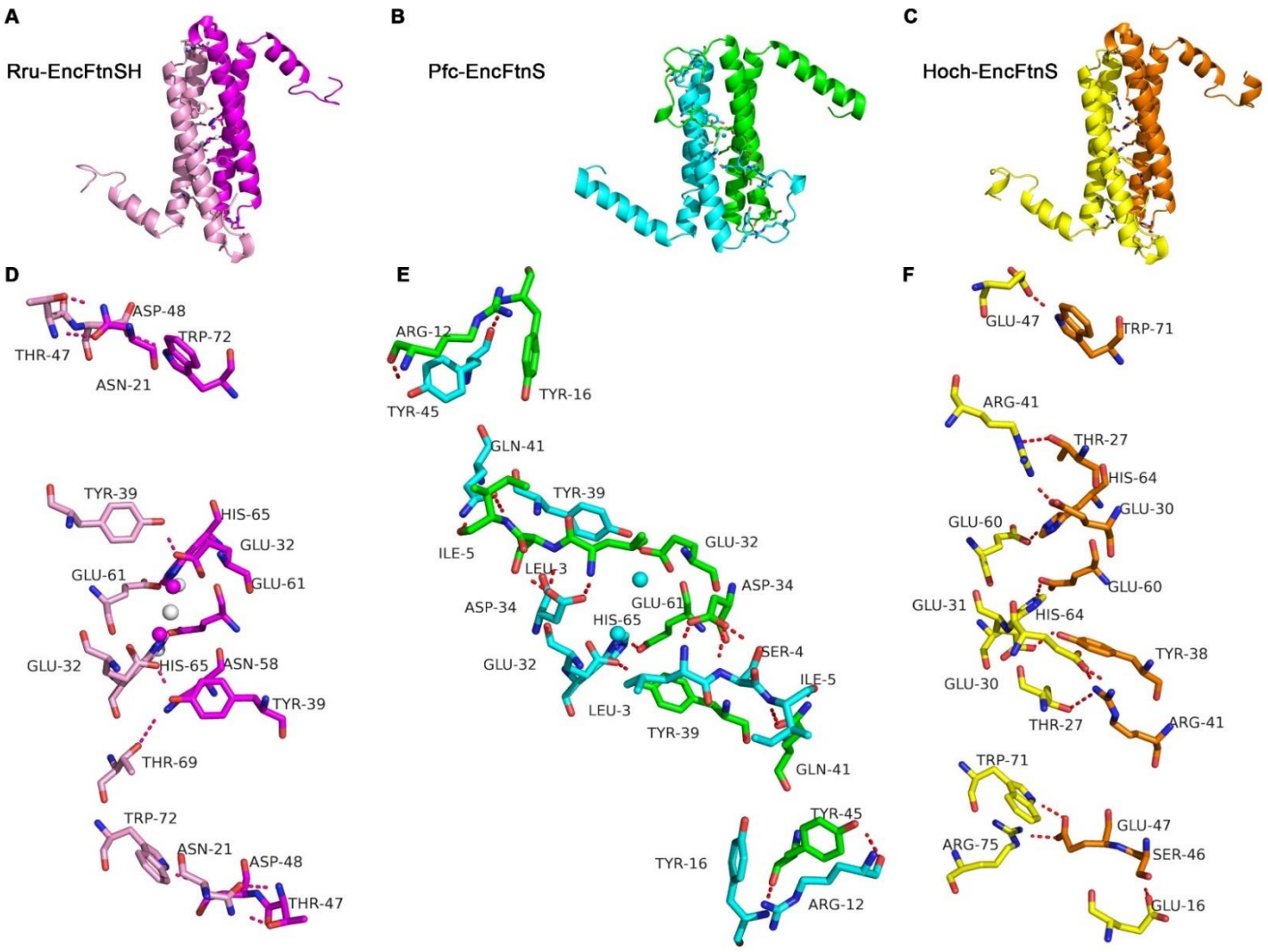


Figure 6.5.2 Polar contacts in the FOC dimer interfaces of EncFtn homologues.

(A/B/C) FOC dimers of Rru-EncFtnSH (pink and purple), Pfc-EncFtnS (cyan and green), Hoch-EncFtnS (yellow and orange) are shown as cartoon representations. (C/ D/E) Residues involved in the polar contacts between FOC dimers of (A/B/C) are shown as sticks. The polar contacts are displayed as red dashed lines between the interacting residues of the two chains.

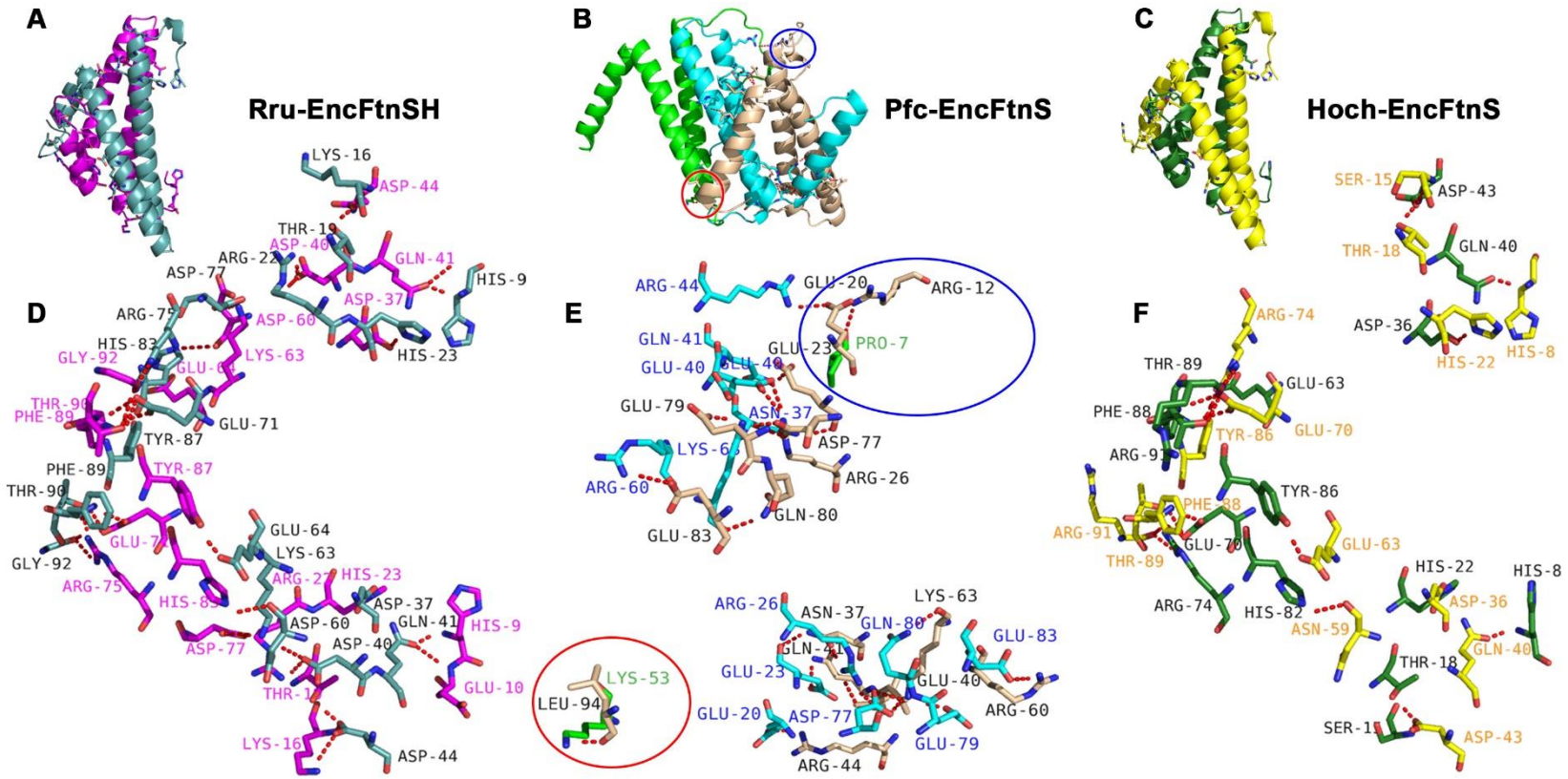


Figure 6.5.3 Polar contacts in the non-FOC dimer interface of EncFtn homologues.

The Non-FOC dimers of (A) Rru-EncFtnSH (purple and light teal), (B) three chains of Pfc-EncFtnS (green, cyan and wheat) and (C) non-FOC dimers of Hoch-EncFtnS (forest and yellow) are shown as cartoon representations. (C/D/E) Residues involved in the polar contacts between non-FOC dimers of (A/B/C) are shown in sticks and labelled with different colours representing 2–3 chains of each protein. The polar contacts are displayed as red dashed lines between the interacting residues of two chains. (B/E) In the Pfc-EncFtnS structure, the long C-terminal α3-helix and the N-terminal loop of the wheat chain are both interacting with the green and cyan chains. The red circles highlight the interacting residues between the C-terminus of the wheat chain and the green chain. The blue circles highlight the interacting residues between the N-terminal loops of the wheat and green chains.

6.6 The putative assembly of *P. furiosus* encapsulin nanocompartment

The structure of the *P. furiosus* encapsulin nanocompartment (KEGG ID: PFC_05175) was determined in the year 2007 and was initially thought to be a virus-like particle (PDB ID: 2E0Z). The native encapsulin purified from *P. furiosus* formed an icosahedron with $T = 3$ symmetry and 180 copies of the 38.8 kDa protein present in the capsid (Akita et al. 2007). The *P. furiosus* encapsulin was made up of 12 pentamers and 20 hexamers. In this article, it was proposed as a single gene with fused *EncFtn* and *encapsulin* encoding a single polypeptide, or EncFtn: Enc = 1:1, which meant that there were 180 subunits of encapsulin and 180 subunits of EncFtn in *P. furiosus* encapsulin nanocompartment. Since the Pfc-EncFtnS was determined as a ring-like decamer structure, the question was how the 12 pentamers and 20 hexamers of encapsulin accommodated 18 copies of decameric EncFtn. The Pfc-EncFtnS decamer structure could only be docked within the pentameric encapsulin structure in analogous to the *R. rubrum* encapsulin nanocompartment model, which remained 20 hexameric encapsulin and 60 subunits of EncFtn. This contradicted with the 1:1 molar ratio of EncFtn and encapsulin in *P. furiosus* encapsulin nanocompartment.

There was no doubt that it was a single gene encoding two fused proteins EncFtn and encapsulin. Firstly, the sequencing of *PFC_05175* gene amplified using *P. furious* DSM 3638 genome as the template confirmed it was a single gene. Secondly, the PFC_05175 protein sequence did not show the C-terminal localisation sequence shared by EncFtn family proteins, indicating that the C-terminus of EncFtn and N-terminus of encapsulin were fused. However, our expression test of *PFC_05175* with optimised codon (pET-28a(+)-optPFC_05175) in *E. coli* BL21(DE3) revealed two major bands around 26 kDa and 38 kDa, which survived 100°C treatment, consistent with the encapsulin protein and the full length of PFC_05175 (**Figure 6.6.1**). The inspection of the RBS of *PFC_05175* mRNA sequence revealed that when expressed in native archaeon *P. furiosus* the first ATG was evaluated as translation rate 4278 au while the GTG at position 327 had the highest translation rate 13540 au, suggesting that an extra translation could have initiated from the GTG (Espah Borujeni et al. 2014; Salis et al. 2009). Interestingly, the sequence translated from GTGGAATATTTTGAG corresponded to VEYFE, which correlated with the N-terminus of PFC_05175

encapsulin structure, indicating that an extra encapsulin protein could be expressed on top of the full-length EncFtn-Enc translated from the first ATG. The deduction, however, was inconsistent with the single band (38 kDa) on SDS-PAGE gel of the encapsulin nanocompartment purified from *P. furiosus* (Namba et al. 2005). When *PFC_05175* was overexpressed in *E. coli* without codon optimisation, the first ATG was assessed as the highest translation rate, indicating that only a single fused EncFtn-encapsulin was produced; theoretically, the encapsulin nanocompartment could not be assembled functionally with 1:1 of EncFtn to encapsulin.

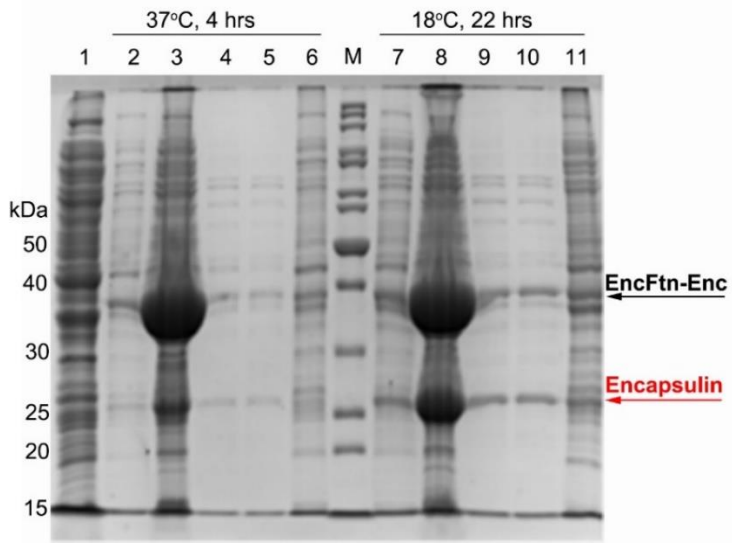


Figure 6.6.1 Expression test of pET-28a(+)-optPFC_05175 in *E. coli* BL21(DE)3 on 15% SDS-PAGE.

E. coli BL21(DE3) harbouring the plasmid pET-28a(+)-optPFC_05175 was expressed in 5 ml LB medium with the induction of 1 mM IPTG under the conditions of 37°C, 4 hrs or 18°C, 22 hrs. Protein marker was loaded in lane M. 1 ml uninduced cells were pelleted by centrifugation at 15,000 *g* for 1 min and loaded into lane 1 as a control. 4 ml induced cells were centrifuged and the pellet was resuspended in 1 ml GF buffer, followed by sonication for 20 s and centrifuged at 15,000 *g* for 1 min. The lysate supernatants were loaded into lanes 2 and 7. The pellet or insoluble proteins were loaded into lanes 3 and 8. The lysate supernatants were heated at 80°C for 10 min, followed by centrifugation; the resulting supernatants were loaded into lanes 4 and 9. The lysate supernatants were heated at 80°C for 10 min and another 100°C for 10 min, followed by centrifugation; the resulting supernatants were loaded into lanes 5 and 10 while the pellets were loaded into lanes 6 and 11. All samples were mixed with 1 x SDS loading buffer, boiled at 100°C for 10 min and centrifuged at 15,000 x *g* for 1 min before loading to each lane. There are two major bands at about 26 kDa and 38 kDa (pointed by black and red arrows) in both supernatant and pellet of the lysate. Both bands are shown in the supernatants of the samples with 80°C or 100°C heat treatment (lanes 4/5 and 9/11), indicating their thermostable property. The 26 kDa band corresponds to the encapsulin protein of PFC_05175 while the 38 kDa band corresponds to the size of PFC_05175.

Taken together, we hypothesised that mRNA sequence of PFC_05175 was translated into two proteins: a full length PFC_05175 protein with fused EncFtn and Enc, and an

extra encapsulin protein of PFC_05175 that formed pentamers (**Figure 6.6.2**). Decameric EncFtn was docked on top of pentameric encapsulin along the 5-fold symmetry axis but not in analogous to the 60-meric encapsulin nanocompartment as mentioned in **Figure 5.7.3**. Instead, the C-termini of EncFtn appeared to connect with the N-termini of hexameric encapsulin nearby. Therefore, it was likely in *P. furiosus* 120 subunits of fused EncFtn-Enc and extra 60 subunits of encapsulin were expressed; EncFtn formed 12 decamers that docked to 12 pentameric encapsulin formed due to the extra translation; and the connecting 120 subunits of encapsulin assembled into 20 hexamers.

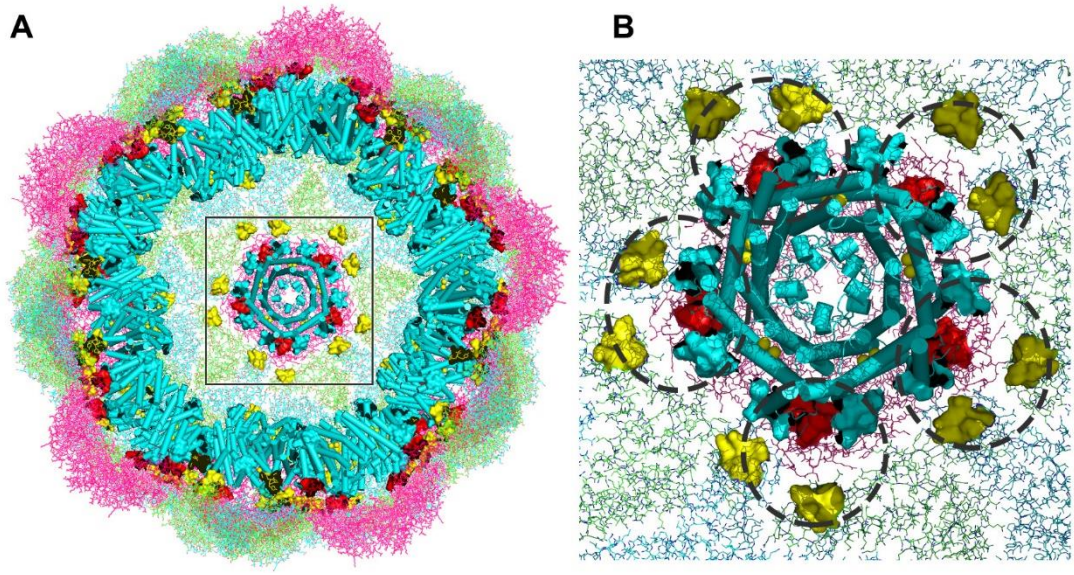


Figure 6.6.2 Hypothesis of *P. furiosus* encapsulin nanocompartment assembly.

(A) *P. furiosus* encapsulin with twelve pentamers in pink lines and twenty hexamers in green/blue lines is fully occupied by twelve decameric EncFtns (cyan cartoons). Each decameric EncFtn is docked on top of a pentameric encapsulin along the 5-fold symmetry axis. (B) Big view of the docking complex highlighted by a black square in (A). The C-termini of EncFtn are shown in blue spheres. The N-termini of pentameric encapsulin are shown in red spheres while the N-termini of hexameric encapsulin are shown in yellow spheres. Five dashed circles highlight the potential connection between the two C-termini of EncFtn with adjacent two N-termini of hexameric encapsulin, correlating with the equivalent expression of EncFtn and encapsulin as a full-length protein. It also suggests that the extra expression of encapsulin assembles into pentamers where EncFtn is superimposed on.

When each pentamer of the *P. furiosus* encapsulin nanocompartment were fully occupied by EncFtn (Pfc-EncFtnS), the inner cavity had a diameter of 20–30 Å, which was much larger than the 8 Å inner diameter of the *T. maritima* encapsulin when it was fully occupied by 12 copies of EncFtn of *R. rubrum*. The larger encapsulin cavity was consistent with the potential for much greater storage and correlating with observations of the *M. xanthus* encapsulin (PDB ID: 4PT2) with 180 subunits, which has been estimated to be able to carry around 30000 iron ions based on the scanning transmission electron microscopy analysis of the dense core inside the encapsulin (McHugh et al. 2014).

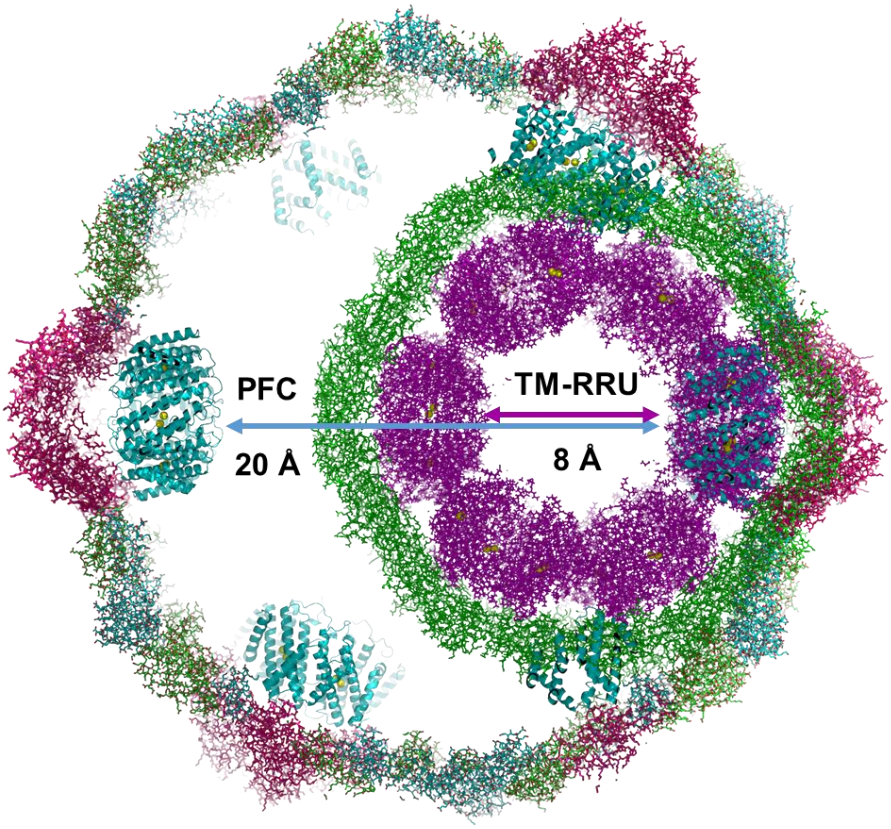


Figure 6.6.3 Interior space comparison of 180-meric and 60-meric encapsulin nanocompartments.

P. furiosus (PFC) encapsulin (the outer circle in lines, PDB ID: 2E0Z) fully occupied with 12 copies of Pfc-EncFtnS (cyan cartoon with iron as yellow spheres) is compared with *T. maritima* encapsulin (TM, inner circle in green lines, PDB ID: 3DKT) fully occupied with 12 copies of decameric Rru-EncFtnSH (RRU, purple lines with iron as yellow spheres) while decameric Pfc-EncFtnS and Rru-EncFtnSH are well aligned. The diameter of the inner space of fully occupied PFC encapsulin nanocompartment is 20–30 Å whereas the inner diameter of fully occupied TM-RRU encapsulin nanocompartment (*T. maritima*-encapsulin docked with *R. rubrum* EncFtn) is 8 Å.

6.7 Conclusions and discussion

In this chapter the structures of EncFtn from the halophilic bacterium *H. ochraceum* and the thermophilic archaeon *P. furiosus* were determined and analysed. The two EncFtn homologues (Hoch-EncFtnS and Pfc-EncFtnS) display a similar annular decamer structure to *R. rubrum* EncFtn (Rru-EncFtnSH).

The Hoch-EncFtnS sequence possessed all same metal binding residues with Rru-EncFtnSH, but no metal was bound in the structure. The metal entry site in Hoch-EncFtnS had a wider space and thus did not coordinate any metal. The metal exit site of Hoch-EncFtnS displayed a similar conformation with Rru-EncFtnSH. Future work involving assay of the ferroxidase activity of Hoch-EncFtnS is necessary to determine whether the protein is indeed active to oxidise Fe(II). Since Hoch-EncFtn was originally from a halobacterium, it is worth exploring if the salt concentration will influence its enzymatic activity. As it was not possible to obtain an iron-bound Hoch-EncFtnS structure, native MS can be conducted to study the behaviour of Hoch-EncFtnS under different pH values with respect to metal binding. ICP-MS can also be applied to explore the metal binding of Hoch-EncFtnS. Further to analysis of the ferroxidase assay of this enzyme, other ferritin-related functions such as peroxidase activity is worth exploring in Hoch-EncFtnS. The assembly and disassembly of Hoch-EncFtnS in the presence of different metal ions and salt concentrations are also worth exploring.

The *P. furiosus* encapsulin nanocompartment (Pfc-EncFtn-Enc, KEGG ID: PFC_05175) was encoded by a single gene with the EncFtn and encapsulin domains produced as a single polypeptide as suggested in the article (Namba et al. 2005), thus the C-terminal localisation sequence shared by EncFtn family proteins was not seen in the EncFtn domain. The Pfc-EncFtnS structure displays a conserved FOC with di-iron coordinated. By contrast, the metal entry site and exit site are not conserved in Pfc-EncFtnS, suggesting a different mechanism for iron movement in the *P. furiosus* encapsulin nanocompartment. Because of the annular decamer assembly of the *P. furiosus* EncFtn, there was no way for 180 subunits of encapsulin to accommodate 180 subunits or EncFtn or 18 decameric EncFtns. Therefore, an assembly model of *P. furiosus* nanocompartment proposed. In contrast to the 60-meric encapsulin

nanocompartment assembly that decameric EncFtn was docked to pentameric encapsulin through the C-terminal extension, the *PFC_05175* mRNA was translated into two proteins: a full length protein with 120 subunits of fused EncFtn and encapsulin forming 20 hexameric encapsulins and 12 decameric EncFtns, and an extra translation of 60 subunits of encapsulin protein assembling into 12 pentamers. The 12 decamers of EncFtn were docked on top of 12 pentameric encapsulins while the C-termini of EncFtn were still connecting with the N-termini of hexameric encapsulin nearby the pentameric encapsulin. To validate this hypothesis, PFC_05175 will be purified and subjected to the native mass spectrometry. Alternatively, the co-expression of Pfc-EncFtn and Pfc-Enc can be performed in *E. coli* using the two multiple cloning sites of pAYCDuet-1 vector, followed by nanocompartments production and analysis by TEM. The size comparison of encapsulin nanocompartments from *P. furiosus* and *T. maritima* revealed that more space in 180-meric encapsulin nanocompartment would allow higher iron storage, and thus could be explored pertaining to the application of iron-loaded encapsulin nanocompartments as magnetic resonance imaging marker.

Chapter 7 Conclusion and outlook

7.1 Conclusion

Encapsulin-associated ferritin-like proteins (EncFtn) represent a unique family class in the ferritin superfamily and are possible to be in the early stage of ferritin evolution. This study provides the first evidence of looking into EncFtn families and understanding the protein assembly, structure and biochemistry of *R. rubrum* EncFtn. The wide distribution of EncFtn in different bacteria or archaea has also been investigated by structural characterisation of EncFtn homologues in *H. ochraceum* and *P. furiosus*.

The recombinant *R. rubrum* EncFtn overexpressed in *E. coli* displays two oligomeric states in solution: monomer and decamer. The ICP-MS result reveals that the monomer contains no detectable iron, while the decamer fractions have a significant amount of iron. The *in vitro* metal titration results reveal that in the gas phase EncFtn decamerises completely in the presence of a five-fold excess of Fe(II). Subsequent analytical gel-filtration and native PAGE results confirm that EncFtn oligomerises in the presence of molar equivalents of Fe(II), Zn(II) or Co(II). Furthermore, Fe(III) does not alter the EncFtn assembly, suggesting the complex has a preference for ferrous iron over ferric iron. Both monomer and decamer fractions are able to crystallise into a ring-doughnut-like decamer structure (5 x 7 x 7 nm) with a central 5-fold symmetry channel at1.6 nm wide, which consists of a pentamer of dimers. The monomer of EncFtn contains two main alpha-helices (α1-α2) and a C-terminal short α3; two monomers arrange in an anti-parallel manner to form two types of dimer, with or without metal associated, named as FOC and non-FOC dimers, respectively. There are three metal binding sites in the FOC dimer interface: metal ion entry site, consisting of E31 and E34 from two chains that coordinate a hydrated calcium ion in the crystal structure; ferroxidase centre (FOC), composed of E32, E62, H65 and Y39 from two chains, which binds di-iron ions; metal ion exit site, comprising H57, E61 and E64 from two chains, which coordinate two calcium ions with partial occupancy in the structure. The single alanine mutations of residues in the FOC abolish the iron binding but do not change the decamer structure of EncFtn. The native mass spectrometry experiment determines the

presence of 15 iron ions in EncFtn, suggesting that metal ion entry/exit site in the structure can be alternative iron binding sites. The alignment of dimeric EncFtn with monomeric bacterial ferritin highlights a common four-helix bundle architecture around the FOC and suggests that classical ferritins may have evolved from EncFtn through gene duplication. Superimposing decameric EncFtn with the 24-meric bacterial ferritin suggests that the two proteins differ in their iron-storage behaviour. In the classical ferritins iron is proposed to move from the outer surface to the FOC where it is oxidised prior to mineralisation in the ferritin nanocage; however, the orientation of the FOC in EncFtn is flipped, suggesting that iron moves from the central channel towards the outer surface of EncFtn. These structural observations lead to major questions: Does EncFtn act as a ferroxidase enzyme? How is iron stored by EncFtn in the absence of an enclosed nanocage structure?

The subsequent ferroxidase assay and end-point ferrozine assay confirm that EncFtn is able to oxidise Fe(II) to Fe(III) and can bind a maximum of 47 iron ions. The ferroxidase activity conferred by the FOC of EncFtn is supported by the diminished iron oxidation of FOC single mutants. The EncFtn gene in *R. rubrum* is in a locus upstream of a gene encoding an encapsulin protein. The recombinantly expressed *R. rubrum* encapsulin protein (Rr-Enc) forms a 24-nm spherical structure with the highest sequence identity to the *T. maritima* encapsulin (Sutter et al. 2008). *R. rubrum* Enc does not possess iron oxidation ability; instead, it binds a maximum of 2200 iron ions, possibly due to a negatively-charged interior surface as seen in the structures of other members of the encapsulin family. The co-expressed Rru-EncFtn and Rru-Enc (Rru-EncFtn-Enc) nanocompartment displays a similar size to Rru-Enc, with an estimated 12 copies of the EncFtn decamer targeted to the interior surface of the pentameric vertices of the encapsulin through the C-terminal location sequence of EncFtn. The EncFtn-Enc complex shows a comparable ferroxidase activity with horse spleen apoferritin. Its maximal iron loading capacity in the assay conditions tested is about 4200 iron ions, which is 7–8 times higher than the iron incorporation in apoferritin in the experiment. Although fully occupied encapsulin nanocompartment shares a similar sized inner cavity with 24-meric apoferritin, at 8 nm diameter, the higher iron loading capacity of EncFtn-Enc suggests a faster iron mineralisation inside the encapsulin cage

due to the 60 FOCs present in the twelve EncFtn decamers in the encapsulin in contrast to the 24 FOCs of classical ferritins.

Based on the protein structure, assembly and biochemical data a model for iron storage can be proposed for the 60-meric encapsulin nanocompartment (EncFtn-Enc). Fe(II) traverses the encapsulin shell through the 5-fold symmetry surface pore, moves along the well-aligned 5-fold central channel of EncFtn, followed by the coordination at the metal ion entry site and translocation to the FOC where Fe(II) is oxidised to Fe(III). Subsequent to oxidation, Fe(III) translocates to the metal ion exit site and is mineralised within the encapsulin cage. The EncFtn-Enc nanocompartment is reminiscent of 24-meric classical ferritins with multiple di-iron centres that may increase the iron oxidation; and the presence of encapsulin is critical for iron mineralisation. Therefore, compared to classical ferritins, the iron oxidation and mineralisation occur in two proteins, with EncFtn oxidising Fe(II) and encapsulin providing the space for the formation of Fe(III) minerals.

The structure and overall arrangement of the EncFtn decamer is conserved across homologues from a halophilic bacterium *H. ochraceum* (Hoch) and a thermophilic archaeon *P. furiosus* (Pfc). There is no metal binding in the Hoch-EncFtn structure, although all metal binding sites are the same as the Rru-EncFtn. Pfc-EncFtn displays more structured N-/C-termini than the other two homologues; this may be related to the fact that the native protein is from a thermophile, which requires stable proteins at high temperatures. The FOC of Pfc-EncFtn is highly conserved, while the metal binding entry/exit sites possess more variable residues without metal association, suggesting differences in the iron oxidation chemistry.

7.2 Outlook

To validate the iron storage model of EncFtn-Enc, it will be necessary to obtain the structure of the Rru-Enc protein. The purification methods used in this work—RNase treatment plus sucrose cushion/gradient ultracentrifugation, produces reasonably pure encapsulin proteins; however, the crystals grown from this protein did not diffract to better than 10 Å. It is possible that the impurities in Rru-Enc have affected the crystal quality. In terms of increasing diffraction quality, it is worth trying to collect

diffraction data at room temperature; alternatively, cryo-protectants can be optimised when collecting diffraction data at 100 K. Besides, different truncated constructs of Rru-Enc can be designed based on cross-linking information, and subjected to crystallisation. Although His-tagged Rru-Enc is easier to purify and gives protein of a much higher purity, it appears to be unstable during dialysis or buffer exchange and precipitates eventually. Other tags such as strep tag can be used to label Rru-Enc, which may help to obtain purer proteins for crystallisation. Alternatively, cryo-EM can also be helpful in determining the Rru-Enc structure.

To understand the transport of Fe(II) ions across the encapsulin shell, molecular dynamic simulations can be performed on models of Rru-Enc to estimate the iron movement across encapsulin shell (Laghaei et al. 2014). Alternatively, hydrogen/deuterium exchange mass spectrometry can be carried out to determine the conformational dynamics of the encapsulin protein (Konermann et al. 2011). Based on these data, surface pore residues of Rru-Enc can be mutated to block the putative entry site and detect if iron oxidation and mineralisation will be affected.

To explore the role of the metal ion entry and exit sites of the Rru-EncFtn in iron oxidation, ferroxidase assays can be performed on metal ion entry/exit mutants. The maximum iron loading capacity can be explored on the corresponding mutations of Rru-EncFtn when enclosed in the encapsulin (Rru-EncFtn-Enc). The conserved tyrosine39, which is proposed to facilitate the iron oxidation as a single electron molecular capacitor in classical ferritins (Ebrahimi, Hagedoorn & Wilfred R Hagen 2013), can be mutated to determine how it affects the iron oxidation of Rru-EncFtn and Rru-EncFtn-Enc.

It is not clear if Hoch-EncFtn and Pfc-EncFtn have iron oxidation ability. It is possible that salt concentration in buffers may affect the assembly and ferroxidase activity of Hoch-EncFtn. Thus ferroxidase assays should be performed on the EncFtn homologues in the future.

The self-assembly of encapsulin nanocompartments as well as the potential for carrying cargos make encapsulins very appealing for biotechnological applications, such as using iron-loaded encapsulin nanocompartments as a medical imaging marker.

References

Abe, S. et al., 2008. Control of the coordination structure of organometallic palladium complexes in an apo-ferritin cage. *Journal of the American Chemical Society*, 130(32), pp.10512–4.

Abramoff, M.D., Magalhaes, P.J. & Ram, S.J., 2004. Image processing with ImageJ. *Biophotonics International*, 11(7), pp.36–42.

Adams, P.D. et al., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.213–221.

Afonine, P. V. et al., 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography*, 68(4), pp.352–367.

Aisen, P., Enns, C. & Wessling-Resnick, M., 2001. Chemistry and biology of eukaryotic iron metabolism. *International Journal of Biochemistry and Cell Biology*, 33(10), pp.940–959.

Akita, F. et al., 2007. The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. *Journal of Molecular Biology*, 368(5), pp.1469–83.

Andrews, S.C. et al., 1995. Site-directed replacement of the coaxial heme ligands of bacterioferritin generates heme-free variants. *Journal of Biological Chemistry*, 270(40), pp.23268–23274.

Andrews, S.C., 2010. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica et Biophysica Acta*, 1800(8), pp.691–705.

Antonyuk, S. V. & Hough, M.A., 2011. Monitoring and validating active site redox states in protein crystals. *Biochimica et Biophysica Acta*, 1814(6), pp.778–784.

Archibald, F., 1983. Lactobacillus plantarum, an organism not requiring iron. *FEMS Microbiology Letters*, 19(1), pp.29–32.

Arnold, A.R. & Barton, J.K., 2013. DNA protection by the bacterial ferritin Dps via DNA charge transport. *Journal of the American Chemical Society*, 135(42), pp.15726–15729.

Ayala, A., Muñoz, M.F. & Argüelles, S., 2014. Lipid peroxidation: Production, metabolism, and signaling mechanisms of malondialdehyde and 4-hydroxy-2-nonenal. *Oxidative Medicine and Cellular Longevity*, 2014.

Bai, L. et al., 2015. Genome-wide comparison of ferritin family from Archaea, Bacteria, Eukarya, and Viruses: its distribution, characteristic motif, and

phylogenetic relationship. *The Science of Nature*, 102(9–10), p.64.

Baker, N.A. et al., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), pp.10037–10041.

Bakkers, G.R. & Boyerg, R.F., 1986. Iron incorporation into apoferritin—The role of apoferritin as a ferroxidase. *The Journal of biological chemistry*, 261(28), pp.13182–13185.

Barreto-Vieira, D.F. & Barth, O.M., 2015. Negative and Positive Staining in Transmission Electron Microscopy for Virus Diagnosis. In *Microbiology in Agriculture and Human Health*. pp. 120–128.

Behera, R.K. & Theil, E.C., 2014. Moving Fe2+ from ferritin ion channels to catalytic OH centers depends on conserved protein cage carboxylates. *Proceedings of the National Academy of Sciences*, 111(22), pp.7925–7930.

Berman, H.M. et al., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28(1), pp.235–242.

Bernacchioni, C. et al., 2014. Loop electrostatics modulates the intersubunit interactions in ferritin. *ACS Chemical Biology*, 9(11), pp.2517–25.

Bertini, I. et al., 2012. Structural insights into the ferroxidase site of ferritins from higher eukaryotes. *Journal of the American Chemical Society*, 134, pp.6169–6176.

Beveridge, R. et al., 2014. A mass-spectrometry-based framework to define the extent of disorder in proteins. *Analytical chemistry*, 86(22), pp.10979–91.

Bonomi, F., Kurtz, D.M. & Cui, X., 1996. Ferroxidase activity of recombinant Desulfovibrio vulgaris rubrerythrin. *Journal of Biological Inorganic Chemistry*, 1(1), pp.67–72.

Bou-abdallah, F. et al., 2002. μ-1,2-peroxobridged di-iron(III) dimer formation in human H-chain ferritin. *Biochemical Journal*, 63, pp.57–63.

Bou-Abdallah, F. et al., 2003. Defining metal ion inhibitor interactions with recombinant human H- and L-chain ferritins and site-directed variants: an isothermal titration calorimetry study. *Journal of Biological Inorganic Chemistry*, 8(4), pp.489–497.

Bou-Abdallah, F. et al., 2014. Functionality of the three-site ferroxidase center of Escherichia coli bacterial ferritin (EcFtnA). *Biochemistry*, 53(3), pp.483–495.

Bou-Abdallah, F. et al., 2005. Origin of the unusual kinetics of iron deposition in human H-chain ferritin. *Journal of the American Chemical Society*, 127(8), pp.3885–3893.

Bou-Abdallah, F. et al., 2004. The putative “nucleation site” in human H-chain ferritin is not required for mineralization of the iron core. *Biochemistry*, 43(14), pp.4332–4337.

Braun, V., 2001. Iron uptake mechanisms and their regulation in pathogenic bacteria. *International Journal of Medical Microbiology*, 291(2), pp.67–79.

Le Brun, N.E. et al., 2010. Iron core mineralisation in prokaryotic ferritins. *Biochimica et Biophysica Acta*, 1800(8), pp.732–44.

Butts, C. a et al., 2008. Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry*, 47(48), pp.12729–39.

Cadet, J. et al., 1999. Hydroxyl radicals and DNA base damage. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 424(1–2), pp.9–21.

Carmona, F. et al., 2013. Ferritin iron uptake and release in the presence of metals and metalloproteins: Chemical implications in the brain. *Coordination Chemistry Reviews*, 257(19–20), pp.2752–2764.

Carpena, X. et al., 2005. A molecular switch and electronic circuit modulate catalase activity in catalase-peroxidases. *EMBO reports*, 6(12), pp.1156–1162.

Cassidy-Amstutz, C. et al., 2016. Identification of a minimal peptide tag for in vivo and in vitro loading of encapsulin. *Biochemistry*, 55, pp.3461–3468.

Ceolín, M., Gálvez, N. & Domínguez-Vera, J.M., 2008. Thermal induced phase transitions and structural relaxation in apoferritin encapsulated copper nanoparticles. *Physical chemistry chemical physics*, 10(29), pp.4327–32.

Chang, C. et al., Crystal Structure of Protein Ne0167 from Nitrosomonas Europaea. *To be Published*.

Charlton, J.R. et al., 2016. Biocompatibility of ferritin-based nanoparticles as targeted MRI contrast agents. *Nanomedicine: Nanotechnology, Biology, and Medicine*, 12(6), pp.1735–1745.

Chasteen, N.D. & Harrison, P.M., 1999. Mineralization in ferritin: an efficient means of iron storage. *Journal of Structural Biology*, 126(3), pp.182–94.

Chen, V.B. et al., 2010. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1), pp.12–21.

Chipperfield, J.R. & Ratledge, C., 2000. Salicylic acid is not a bacterial siderophore: A theoretical study. *BioMetals*, 13(2), pp.165–168.

Cho, K.J. et al., 2009. The crystal structure of ferritin from Helicobacter pylori reveals unusual conformational changes for iron uptake. *Journal of Molecular Biology*, 390(1), pp.83–98.

Colombo, C. et al., 2014. Review on iron availability in soil: Interaction of Fe minerals, plants, and microbes. *Journal of Soils and Sediments*, 14(3), pp.538–548.

Contreras, H. et al., 2014. Characterization of a Mycobacterium tuberculosis nanocompartment and its potential cargo proteins. *The Journal of Biological Chemistry*, 289(26), pp.18279–18289.

Crichton, R., 2009. The importance of iron for biological systems. In *Iron metabolism*. John Wiley & Sons, Ltd, pp. 17–58.

Crichton, R.R. & Declercq, J.-P., 2010. X-ray structures of ferritins and related proteins. *Biochimica et Biophysica Acta*, 1800(8), pp.706–718.

Crooks, G.E. et al., 2004. WebLogo: A sequence logo generator. *Genome Research*, 14(6), pp.1188–1190.

Crow, A. et al., 2009. Structural basis for iron mineralization by bacterioferritin. *Journal of the American Chemical Society*, 131(19), pp.6808–6813.

Cutler, C. et al., 2005. Iron loading into ferritin can be stimulated or inhibited by the presence of cations and anions: a specific role for phosphate. *Journal of Iorganic Biochemistry*, 99(12), pp.2270–2275.

Dautant, A. et al., 1998. Structure of a monoclinic crystal form of Cytochrome b1 (Bacterioferritin) from E. coli. *Acta Crystallographica Section D: Biological Crystallography*, 54(1), pp.16–24.

Davies, K.J., 1987. Protein damage and degradation by oxygen radicals. I. general aspects. *Journal of Biological Chemistry*, 262(20), pp.9895–9901.

Deans, A.E. et al., 2006. Cellular MRI contrast via coexpression of transferrin receptor and ferritin. *Magnetic Resonance in Medicine*, 56(1), pp.51–59.

DeMaré, F., Kurtz, D.M. & Nordlund, J. and P., 1996. The structure of Desulfovibrio vulgaris rubrerythrin reveals a unique combination of rubredoxin-like FeS4 and ferritin-like diiron domains. *Nature Structural & Molecular Biology*, 3, pp.539–546.

Douglas, T. & Ripoll, D.R., 1998. Calculated electrostatic gradients in recombinant human H-chain ferritin. *Protein Science*, 7(5), pp.1083–1091.

Ebrahimi, K.H. et al., 2009. Catalysis of iron core formation in Pyrococcus furiosus ferritin. *Journal of Biological Inorganic Chemistry*, 14(8), pp.1265–74.

Ebrahimi, K.H. et al., 2012. The catalytic center of ferritin regulates iron storage via Fe(II)-Fe(III) displacement. *Nature Chemical Biology*, 8, pp.941–948.

Ebrahimi, K.H., Hagedoorn, P.-L. & Hagen, W.R., 2013. A conserved tyrosine in ferritin is a molecular capacitor. *Chembiochem*, 14(9), pp.1123–33.

Ebrahimi, K.H., Hagedoorn, P.-L. & Hagen, W.R., 2010. Inhibition and stimulation of formation of the ferroxidase center and the iron core in Pyrococcus furiosus ferritin. *Journal of Biological Inorganic Chemistry*, 15(8), pp.1243–53.

Ebrahimi, K.H., Hagedoorn, P.-L. & Hagen, W.R., 2013. Phosphate accelerates displacement of Fe(III) by Fe(II) in the ferroxidase center of Pyrococcus furiosus ferritin. *FEBS Letters*, 587(2), pp.220–225.

Ebrahimi, K.H., Hagedoorn, P.-L. & Hagen, W.R., 2015a. Self-assembly is prerequisite for catalysis of Fe(II) oxidation by catalytically active subunits of ferritin. *The Journal of Biological Chemistry*, 290, pp.26801–26810.

Ebrahimi, K.H., Hagedoorn, P.-L. & Hagen, W.R., 2015b. Unity in the Biochemistry of the Iron-Storage Proteins Ferritin and Bacterioferritin. *Chemical Reviews*, 115(1), pp.295–326.

Emsley, P. et al., 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp.486–501.

Espah Borujeni, A., Channarasappa, A.S. & Salis, H.M., 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4), pp.2646–2659.

Evans, P.R., 2011. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp.282–292.

Evans, P.R. & Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography*, 69(7), pp.1204–1214.

Fan, J. et al., 2011. Direct evidence for catalase and peroxidase activities of ferritin-platinum nanoparticles. *Biomaterials*, 32(6), pp.1611–8.

Fan, R. et al., 2010. Fabrication of gold nanoparticles inside unmodified horse spleen apoferritin. *Small*, 6(14), pp.1483–1487.

Fenton, H.J.H., 1876. On a new reaction of tartaric acid. *Chem. News*, 33, p.190.

Fernandez de la Mora, J., 2000. Electrospray ionization of large multiply charged species proceeds via Dole ' s charged residue mechanism. *Analytical Chimica Acta*, 406(1), pp.93–104.

Folta-Stogniew, E., 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods in Molecular Biology*, 328, pp.97–112.

Ford, G.C. et al., 1984. Ferritin: design and formation of an iron-storage molecule. *Philosophical transactions of the Royal Society of London. Series B, Biological*

sciences, 304, pp.551–565.

Frottin, F. et al., 2006. The Proteomics of N-terminal Methionine Cleavage. *Molecular & Cellular Proteomics*, 5(12), pp.2336–2349.

Fudou, R. et al., 2002. Haliangium ochraceum gen. nov., sp. nov. and Haliangium tepidum sp. nov.: novel moderately halophilic myxobacteria isolated from coastal saline environments. *The Journal of General and Applied Microbiology*, 48(2), pp.109–116.

Gallois, B. et al., 1997. X-ray structure of recombinant horse L-chain apoferritin at 2.0 Å resolution: implications for stability and function. *Journal of Biological Inorganic Chemistry*, 2(3), pp.360–367.

Galvez, N. et al., 2008. Comparative Structural and Chemical Studies of Ferritin. *Journal of the American Chemical Society*, 130(5), pp.8062–8068.

Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*. pp. 571–607.

Giessen, T.W. & Silver, P.A., 2016. Microbes use encapsulin protein organelles to sequester toxic reactions. *bioRxiv*.

Granier, T. et al., 2003. Structural description of the active sites of mouse L-chain ferritin at 1.2 Å resolution. *Journal of Biological Inorganic Chemistry*, 8(1), pp.105–111.

Grant, R.A. et al., 1998. The crystal structure of Dps, a ferritin homolog that binds and proteins DNA. *Nature Structural Biology*, 5, pp.294–303.

Gursahani, S., Schoephoerster, R.T. & Prabhakaran, M., 2008. Exploring electron transfer between heme proteins of cytochrome C super family in biosensors: a molecular mechanics study. *Journal of Biomolecular Structure & Dynamics*, 26(3), pp.329–338.

Ha, Y. et al., 1999. Crystal structure of bullfrog M ferritin at 2.8 Å resolution: analysis of subunit interactions and the binuclear metal center. *Journal of Biological Inorganic Chemistry*, 4, pp.243–256.

Hall, T.A., 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, (41), pp.95–98.

Hall, Z. et al., 2013. The role of salt bridges, charge density, and subunit flexibility in determining disassembly routes of protein complexes. *Structure*, 21(8), pp.1325–1337.

Harding, M.M., Nowicki, M.W. & Walkinshaw, M.D., 2010. Metals in protein structures: a review of their principal features. *Crystallography Reviews*, 16(4), pp.247–302.

He, D. & Marles-Wright, J., 2015. Ferritin family proteins and their use in bionanotechnology. *New Biotechnology*, 32(6), pp.651–657.

Hempstead, P.D. et al., 1997. Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *Journal of Molecular Biology*, 268(2), pp.424–448.

Hendrickson, W.A., Horton, J.R. & LeMaster, D.M., 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal*, 9(5), pp.1665–1672.

Hicks, P.M. et al., 1998. Homomultimeric protease in the hyperthermophilic bacterium Thermotoga maritima has structural and amino acid sequence homology to bacteriocins in mesophilic bacteria. *FEBS Letters*, 440(3), pp.393–398.

Huard, D.J.E., Kane, K.M. & Tezcan, F.A., 2013. Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nature Chemical Bology*, 9(3), pp.169–76.

Huber, R. et al., 1986. Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Archives of Microbiology*, 144(4), pp.324–333.

Hudson, A.J. et al., 1993. Overproduction, purification and characterization of the Escherichia coli ferritin. *European Journal of Biochemistry*, 218(3), pp.985–95.

Iverson, S. et al., 2016. CIDAR MoClo: improved MoClo assembly standard and new E. coli part library enables rapid combinatorial design for synthetic and traditional biology. *ACS Synthetic Biology*, 5, pp.59–103.

Jacobs, J.F. et al., 2010. Development of a bionanotechnological phosphate removal system with thermostable ferritin. *Biotechnology and Bioengineering*, 105(5), pp.918–923.

Jameson, G.N.L. et al., 2002. Stoichiometric production of hydrogen peroxide and parallel formation of ferric multimers through decay of the diferric-peroxo complex, the first detectable intermediate in ferritin mineralization. *Biochemistry*, 41(45), pp.13435–13443.

Jin, R. et al., 2014. Superparamagnetic iron oxide nanoparticles for MR imaging and therapy: design considerations and clinical applications. *Current opinion in pharmacology*, 18C, pp.18–27.

Johnson, L., 2008. Iron and siderophores in fungal-host interactions. *Mycological Research*, 112(2), pp.170–183.

Jones, T., Spencer, R. & Walsh, C., 1978. Mechanism and kinetics of iron release from ferritin by dihydroflavins and dihydroflavin analogues. *Biochemistry*, 17(19), pp.4011–4017.

Kabsch, W., 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.133–144.

Kanekiyo, M. et al., 2015. Rational design of an Epstein-Barr virus vaccine targeting the receptor-binding site. *Cell*, 162, pp.1090–1100.

Kanekiyo, M. et al., 2013. Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature*, 499(7456), pp.102–6.

Kanu, A.B. et al., 2008. Ion mobility-mass spectrometry. *Journal of Mass Spectrometry*, 43(1), pp.1–22.

Karplus, P.A. & Diederichs, K., 2012. Linking crystallographic model and data quality. *Science*, 336(6084), pp.1030–1033.

Kasyutich, O. et al., 2010. Silver ion incorporation and nanoparticle formation inside the cavity of Pyrococcus furiosus ferritin: structural and size-distribution analyses. *Journal of the American Chemical Society*, 132(10), pp.3621–7.

Katoh, K. et al., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), pp.3059–3066.

Katz, A.K. et al., 1996. Calcium ion coordination: A comparison with that of beryllium, magnesium, and zinc. *Journal of the American Chemical Society*, 118(24), pp.5752–5763.

Kidane, T.Z., Sauble, E. & Linder, M.C., 2006. Release of iron from ferritin requires lysosomal activity. *American Journal of Physiology. Cell Physiology*, 291, pp.C445–C455.

Kim, M. et al., 2011. pH-Dependent Structures of Ferritin and Apoferritin in Solution: Disassembly and Reassembly. *Biomacromolecules*, 12, pp.1629–1640.

Klem, M.T. et al., 2008. Photochemical mineralization of europium, titanium, and iron oxyhydroxide nanoparticles in the ferritin protein cage. *Inorganic chemistry*, 47(7), pp.2237–2239.

Konermann, L., Pan, J. & Liu, Y.-H., 2011. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society reviews*, 40(3), pp.1224–1234.

Konijnenberg, A., Butterer, A. & Sobott, F., 2013. Native ion mobility-mass spectrometry and related methods in structural biology. *Biochimica et Biophysica Acta*, 1834(6), pp.1239–56.

Krissinel, E. & Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3), pp.774–797.

Krissinel, E. & Henrick, K., 2004. Secondary-structure matching (SSM), a new tool

for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D: Biological Crystallography*, 60, pp.2256–2268.

Kumar, S., Stecher, G. & Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*.

Kümmerle, R. et al., 2000. Electron transfer properties of iron-sulfur proteins. *Journal of inorganic biochemistry*, 79(1–4), pp.83–91.

Laghaei, R. et al., 2014. Calculation of iron transport through human H-chain ferritin. *The Journal of Physical Chemistry*.

Langlois d'Estaintot, B. et al., 2004. Crystal structure and biochemical properties of the human mitochondrial ferritin and its mutant Ser144Ala. *Journal of Molecular Biology*, 340, pp.277–293.

Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947–2948.

Laufberger, M., 1937. Sur la cristallisation de la ferritine. *Bulletin de la Societe de Chimie Biologique*, 19, pp.1575–1582.

Laulhère, J.P., Barcelò, F. & Fontecave, M., 1995. Dynamic equilibria in iron uptake and release by ferritin. *BioMetals*, 9, pp.303–309.

Lawson, D.M. et al., 1991. Solving the structure of human H-ferritin by genetically engineering intermolecular crystal contacts. *Nature*, 349, pp.541–544.

Levi, S. et al., 1988. Mechanism of ferritin iron uptake: activity of the H-chain and deletion mapping of the ferro-oxidase site. A study of iron uptake and ferro-oxidase activity of human liver, recombinant H-chain ferritins, and of two H-chain deletion mutants. *Journal of Biological Chemistry*, 263(34), pp.18086–18092.

Levi, S. et al., 1994. The role of the L-chain in ferritin iron incorporation. *Journal of Molecular Biology*, 238(5), pp.649–654.

Li, W. et al., 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, 43(W1), pp.W580-4.

Lill, R., 2009. Function and biogenesis of iron-sulphur proteins. *Nature*, 460(7257), pp.831–8.

Listowsky, I. et al., 1972. Denaturation of horse spleen ferritin in aqueous guanidinium chloride solutions. *Biochemistry*, 11(11), pp.2176–2182.

Macedo, S. et al., 2003. The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. *Nature Structural Biology*, 10(4), pp.285–90.

MacKenzie, E.L., Iwasaki, K. & Tsuji, Y., 2008. Intracellular iron transport and

storage: from molecular mechanisms to health implications. *Antioxidants & Redox Signaling*, 10(6), pp.997–1030.

Mann, S., Bannister, J. V. & Williams, R.J.P., 1986. Structure and composition of ferritin cores isolated from human spleen, limpet (Patella vulgata) hemolymph and bacterial (Pseudomonas aeruginosa) cells. *Journal of Molecular Biology*, 188(2), pp.225–232.

Marklund, E.G. et al., 2015. Collision cross sections for structural proteomics. *Structure*, 23(4), pp.791–799.

Masuda, T. et al., 2010a. Crystal structure of plant ferritin reveals a novel metal binding site that functions as a transit site for metal transfer in ferritin. *The Journal of Biological Chemistry*, 285(6), pp.4049–4059.

Masuda, T. et al., 2010b. The universal mechanism for iron translocation to the ferroxidase site in ferritin, which is mediated by the well conserved transit site. *Biochemical and Biophysical Research Communications*, 400(1), pp.94–9.

McCoy, A.J. et al., 2007. Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), pp.658–674.

McHugh, C.A. et al., 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO journal*, 33(14), pp.1–16.

McWilliam, H. et al., 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic acids research*, 41(Web Server issue).

Mertz, J.R. & Theil, E.C., 1983. Subunit dimers in sheep spleen apoferritin. The effect on iron storage. *Journal of Biological Chemistry*, 258(19), pp.11719–11726.

Moon, H. et al., 2014. Genetically engineering encapsulin protein cage nanoparticle as a SCC-7 cell targeting optical nanoprobe. *Biomaterials Research*, 18(21), pp.1–7.

Namba, K. et al., 2005. Expression and molecular characterization of spherical particles derived from the genome of the hyperthermophilic euryarchaeote Pyrococcus furiosus. *Journal of Biochemistry*, 138(2), pp.193–199.

Nei, M. & Kumar, S., 2000. *Molecular evolution and phylogenetics*, Oxford university press.

Newman, J. et al., 2005. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D*, 61(10), pp.1426–1431.

Notredame, C., Higgins, D.G. & Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, 302(1), pp.205–17.

Ogata, H. et al., 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1), pp.29–34.

Pannu, N.S. et al., 1998. Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallographica Section D: Biological Crystallography*, 54(6 II), pp.1285–1294.

Pei, J., Kim, B.-H. & Grishin, N. V, 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research*, 36(7), pp.2295–300.

Pereira, A.S. et al., 1998. Direct spectroscopic and kinetic evidence for the involvement of a peroxodiferric intermediate during the ferroxidase reaction in fast ferritin mineralization. *Biochemistry*, 37(28), pp.9871–9876.

Pesek, J. et al., 2011. Structure and mechanism of iron translocation by a Dps protein from Microbacterium arborescens. *The Journal of Biological Chemistry*, 286(40), pp.34872–34882.

Pfaffen, S. et al., 2013. Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. *The Journal of Biological Chemistry*, 288(21), pp.14917–25.

Polanams, J., Ray, A.D. & Watt, R.K., 2005. Nanophase iron phosphate, iron arsenate, iron vanadate, and iron molybdate minerals synthesized within the protein cage of ferritin. *Inorganic Chemistry*, 44(9), pp.3203–9.

Ponka, P., 1999. Cell biology of heme. *The American Journal of the Medical Sciences*, 318(4), pp.241–256.

Pozzi, C. et al., 2015. Iron binding to human heavy-chain ferritin. *Acta Crystallographica Section D: Biological Crystallography*, 71, pp.1909–1920.

Proudhon, D. et al., 1996. Ferritin gene organization: differences between plants and animals suggest possible Kingdom-specific selective constraints. *Journal of Molecular Evolution*, 42, pp.325–336.

Rahmanpour, R. & Bugg, T.D.H., 2013. Assembly in vitro of Rhodococcus jostii RHA1 encapsulin and peroxidase DypB to form a nanocompartment. *FEBS Journal*, 280(9), pp.2097–2104.

Ratledge, C. & Dover, L.G., 2000. Iron metabolism in pathogenic bacteria. *Annual Review of Microbiology*, 54, pp.881–941.

Reed, C.J. et al., 2013. Protein adaptations in archaeal extremophiles. *Archaea*, 2013.

Rice, P., Longden, I. & Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(1), pp.276–277.

Riemer, J. et al., 2004. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. *Analytical Biochemistry*, 331(2), pp.370–5.

Riley, P.A., 1994. Free Radicals in Biology: Oxidative Stress and the Effects of Ionizing Radiation. *International Journal of Radiation Biology*, 65(1), pp.27–33.

Robert, X. & Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1).

Rohrer, J.S. et al., 1990. Iron environment in ferritin with large amounts of phosphate, from Azotobacter vinelandii and horse spleen, analyzed using extended X-ray absorption fine structure (EXAFS). *Biochemistry*, 29(1), pp.259–264.

Rosenkrands, I. et al., 1998. Identification and characterization of a 29-kilodalton protein from Mycobacterium tuberculosis culture filtrate recognized by mouse memory effector cells. *Infection and Immunity*, 66(6), pp.2728–2735.

Rucker, P., Torti, F.M. & Torti, S. V, 1996. Role of H and L subunits in mouse Fferritin. *Journal of Biological Chemistry*, 271(52), pp.33352–33357.

Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), pp.406–25.

Salis, H.M., Mirsky, E.A. & Voigt, C.A., 2009. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, 27(10), pp.946–50.

Sandy, M. & Butler, A., 2009. Microbial Iron Acquisition: Marine and Terrestrial Siderophores. *Chemical Reviews*, 109(10), pp.4580–4595.

Schrodinger, L., 2015. The PyMOL Molecular Graphics System, Version 1.8.

Sheng, Y. et al., 2014. Superoxide dismutases and superoxide reductases. *Chemical Reviews*, 114(7), pp.3854–3918.

Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), p.539.

Siglioccolo, A. et al., 2011. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC Sructural Biology*, 11, p.50.

de Silva, D., Guo, J.-. H. & Aust, S.D., 1993. Relationship between iron and phosphate in mammalian ferritins. *Archives of Biochemistry and Biophysics*, 303(2), pp.451–455.

Simonian, M.H. & Smith, J.A., 2001. Spectrophotometric and Colorimetric Determination of Protein Concentration. In *Current Protocols in Molecular Biology*. p. 76:I:10.1A:10.1.1–10.1A.9.

Snijder, J. et al., 2016. Assembly and mechanical properties of the cargo-free and

cargo-loaded bacterial nanocompartment encapsulin. *Biomacromolecules*, 17, pp.2522–2529.

Snijder, J. et al., 2014. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *Journal of the American Chemical Society*, 136(20), pp.7295–9.

Solovyev, V. V & Salamov, A., 2011. Automatic annotation of microbial genomes and metagenomic sequences. In R. W. Li, ed. *Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers.

Stein, N., 2008. CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography*, 41(3), pp.641–643.

Steiner, R.A., Lebedev, A.A. & Murshudov, G.N., 2003. Fisher's information in maximum-likelihood macromolecular crystallographic refinement. *Acta Crystallographica - Section D Biological Crystallography*, 59(12), pp.2114–2124.

Sun, S. et al., 1993. Ferroxidase kinetics of human liver apoferritin, recombinant H-chain apoferritin, and site-directed mutants. *Biochemistry.*, 32(36), pp.9362–9369.

Sutter, M. et al., 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology*, 15(9), pp.939–947.

Sutter, M., 2008. *Structural basis of enzyme encapsulation into a bacterial nanocompartment*. ETH Zurich.

Svensson-Ek, M. et al., 2002. The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from Rhodobacter sphaeroides. *Journal of Molecular Biology*, 321(2), pp.329–339.

Swartz, L. et al., 2006. Redox-dependent structural changes in the Azotobacter vinelandii bacterioferritin: New insights into the ferroxidase and iron transport mechanism. *Biochemistry*, 45(14), pp.4421–4428.

Takahashi, T. & Kuyucak, S., 2003. Functional properties of threefold and fourfold channels in ferritin deduced from electrostatic calculations. *Biophysical Journal*, 84(4), pp.2256–63.

Tamura, A. et al., 2014. Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnology and Bioengineering*, pp.1–29.

Tatur, J., Hagen, W.R. & Matias, P.M., 2007. Crystal structure of the ferritin from the hyperthermophilic archaeal anaerobe Pyrococcus furiosus. *Journal of Biological Inorganic Chemistry*, 12(5), pp.615–30.

Theil, E. & Raymond, K., 1994. Transition-metal storage, transport, and biomineralization. In *Bioinorganic Chemistry*. pp. 1–35.

Theil, E.C., 2003. Ferritin : At the Crossroads of Iron and Oxygen Metabolism. *Journal of Nutrition*, 133(5), pp.1549–1553.

Theil, E.C., 1987. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. *Annual Review of Biochemistry*, 56, pp.289–315.

Theil, E.C. et al., 2000. The ferritin iron entry and exit problem. *Inorganica Chimica Acta*, 297(1–2), pp.242–251.

Theil, E.C., Liu, X.S. & Tosha, T., 2008. Gated pores in the ferritin protein nanocage. *Inorganica Chimica Acta*, 361(4), pp.868–874.

Thiede, B. et al., 2005. Peptide mass fingerprinting. *Methods*, 35(3 SPEC.ISS.), pp.237–247.

Thore, A., Keister, D.L. & Pietro, A.S., 1969. Studies on the respiratory system of aerobically (dark) and anaerobically (light) grown Rhodospirillum rubrum. *Arch. Mikrobiol.*, 67, pp.378–396.

Tosha, T. et al., 2010. Moving metal ions through ferritin-protein nanocage from three-fold pores to catalytic sites. *Journal of the American Chemical Society*, 132(41), pp.14562–9.

Toussaint, L. et al., 2009. Comparative Fe and Zn K-edge X-ray absorption spectroscopic study of the ferroxidase centres of human H-chain ferritin and bacterioferritin from Desulfovibrio desulfuricans. *Journal of Biological Inorganic Chemistry*, 14(1), pp.35–49.

Trefry, a, Harrison, P.M. & Treffry, B.A., 1978. Incorporation and release of inorganic phosphate in horse spleen ferritin. *The Biochemical Journal*, 171(2), pp.313–20.

UniProt Consortium, 2014. UniProt: a hub for protein information. *Nucleic Acids Research*, (43), pp.D204-12.

Valdés-Stauber, N. & Scherer, S., 1994. Isolation and characterization of Linocin M18, a bacteriocin produced by Brevibacterium linens. *Applied and Environmental Microbiology*, 60(10), pp.3809–3814.

Waldron, K.J. et al., 2009. Metalloproteins and metal sensing. *Nature*, 460, pp.823–30.

Waldron, K.J. & Robinson, N.J., 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews. Microbiology*, 7(1), pp.25–35.

Wang, Q., Mercogliano, C.P. & Löwe, J., 2011. A ferritin-based label for cellular electron cryotomography. *Structure*, 19(2), pp.147–54.

Wang, Z. et al., 2006. Structure of human ferritin L chain. *Acta crystallographica. Section D, Biological crystallography*, 62, pp.800–806.

Wardman, P. & Candeias, L.P., 1996. Fenton chemistry: an introduction. *Radiation Research*, 145(5), pp.523–531.

Watt, G.D. et al., 1992. Fe2+ and phosphate interactions in bacterial ferritin from Azotobacter vinelandii. *Biochemistry*, 31(24), pp.5672–5679.

Watt, G.D. & Jacobst, D., 1988. Redox reactivity of bacterial and mammalian ferritin: Is reductant entry into the ferritin interior a necessary step for iron release? *Proceedings of the National Academy of Sciences*, 85, pp.7457–7461.

Watt, R.K., Hilton, R.J. & Graff, D.M., 2010. Oxido-reduction is not the only mechanism allowing ions to traverse the ferritin protein shell. *Biochimica et Biophysica Acta*, 1800(8), pp.745–59.

Weeratunga, S.K. et al., 2009. Binding of Pseudomonas aeruginosa apobacterioferritin-associated ferredoxin to bacterioferritin B promotes heme mediation of electron delivery and mobilization of core mineral iron. *Biochemistry*, 48(31), pp.7420–31.

Weeratunga, S.K. et al., 2010. Structural studies of bacterioferritin B from Pseudomonas aeruginosa suggest a gating mechanism for iron uptake via the ferroxidase center. *Biochemistry*, 49(6), pp.1160–1175.

Weinberg, M. & Jenney, F., 2004. Rubrerythrin from the hyperthermophilic archaeon Pyrococcus furiosus is a rubredoxin-dependent, iron-containing peroxidase. *Journal of Bacteriology*, 186(23), pp.7888–7895.

Wong, S.G. et al., 2014. The B-type Channel is a Major Route for Iron Entry into the Ferroxidase Center and Central Cavity of Bacterioferritin. *The Journal of Biological Chemistry*, 290, pp.3732–3739.

Yang, X. et al., 1998. Reaction paths of iron oxidation and hydrolysis in horse spleen and recombinant human ferritins. *Biochemistry*, 37(27), pp.9743–9750.

Yang, X. et al., 2000. The iron oxidation and hydrolysis chemistry of Escherichia coli bacterioferritin. *Biochemistry*, 39(16), pp.4915–4923.

Yang, X., Arosio, P. & Chasteen, N.D., 2000. Molecular diffusion into ferritin: pathways, temperature dependence, incubation time, and concentration effects. *Biophysical Journal*, 78(4), pp.2049–2059.

Yang, X. & Chasteen, N.D., 1996. Molecular diffusion into horse spleen ferritin: a nitroxide radical spin probe study. *Biophysical Journal*, 71(3), pp.1587–1595.

Yasmin, S. et al., 2011. A new role for heme, facilitating release of iron from the bacterioferritin iron biomineral. *The Journal of Biological Chemistry*, 286(5), pp.3473–3483.

Yoshimura, H., 2006. Protein-assisted nanoparticle synthesis. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 282–283, pp.464–470.

Zámocký, M. et al., 2015. Independent evolution of four heme peroxidase superfamilies. *Archives of Biochemistry and Biophysics*, 574, pp.108–119.

Zeth, K., 2012. Dps biomineralizing proteins: multifunctional architects of nature. *The Biochemical Journal*, 445(3), pp.297–311.

Zeth, K., Hoiczyk, E. & Okuda, M., 2016. Ferroxidase-Mediated Iron Oxide Biomineralization: Novel Pathways to Multifunctional Nanoparticles. *Trends in Biochemical Sciences*, 41(2), pp.190–203.

Zhang, Y. et al., 2010. Alanine-shaving mutagenesis to determine key interfacial residues governing the assembly of a nano-cage maxi-ferritin. *Journal of Biological Chemistry*, 285(16), pp.12078–12086.

Zhang, Y. & Orner, B.P., 2011. Self-assembly in the ferritin nano-cage protein superfamily. *International Journal of Molecular Sciences*, 12, pp.5406–5421.

Zhao, G. et al., 2001. Is hydrogen peroxide produced during iron(II) oxidation in mammalian apoferritins. *Biochemistry*, 40(36), pp.10832–10838.

Zhen, Z. et al., 2013. RGD-modified apoferritin nanoparticles for efficient drug delivery to tumors. *ACS Nano*, 7(6), pp.4830–4837.

Zheng, H. et al., 2014. Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nature Protocols*, 9(1), pp.156–170.

Van Zundert, G.C.P. et al., 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, 428(4), pp.720–725.

Appendices

Protein name	Protein sequence	Number of amino acids	Predicted MW (Da)	Predicted pI	Extinction coefficient (M^{-1} cm^{-1})	Abs 0.1% (=1 g/l)
Rru-EncFtn	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAIEAADTAGEGSGGDAAKGATAQGD GSLGIGSLKGEAALARPPRL	140	15192.74	4.75	19480	1.282
Rru-EncFtnSH	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13325.67	5.34	19480	1.462
Rru-EncFtnSH-E32A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEALEAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13267.63	5.34	19480	1.468

Rru-EncFtnSH-E62A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDWYD QRVDASTDPELTAILAHNRDEAKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13267.63	5.45	19480	1.468
Rru-EncFtnSH-H65A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEALEAVDWYD QRVDASTDPELTAILAHNRDEEKEAAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13201.57	5.35	19480	1.476
Rru-EncFtnSH-E32D	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEDLEAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13311.64	5.33	19480	1.463
Rru-EncFtnSH-E62D	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDWYD QRVDASTDPELTAILAHNRDEDKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13311.64	5.33	19480	1.463
Rru-EncFtnSH-E32A-E62A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEALEAVDWYD QRVDASTDPELTAILAHNRDEAKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13209.59	5.57	19480	1.475

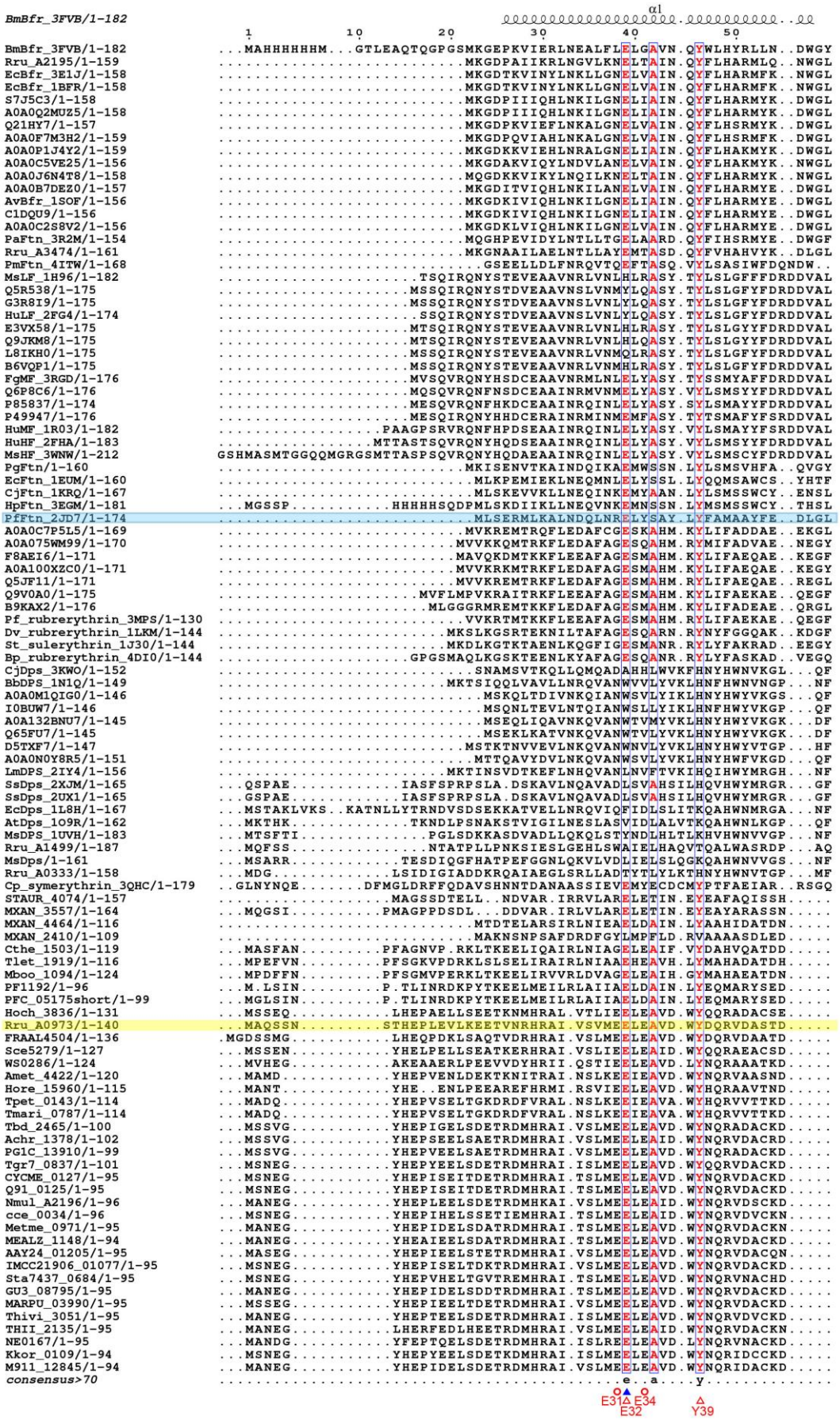
Rru-EncFtnSH-E32D-E62D	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEDLEAVDWYD QRVDASTDPELTAILAHNRDEDKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13297.61	5.32	19480	1.465
Rru-EncFtnSH-E32A-E62A-H65A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEALEAVDWYD QRVDASTDPELTAILAHNRDEAKEAAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13143.53	5.47	19480	1.482
Rru-EncFtnSH-E31A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMAELEAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13267.63	5.45	19480	1.468
Rru-EncFtnSH-E34A	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELAAVDWYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13267.63	5.45	19480	1.468

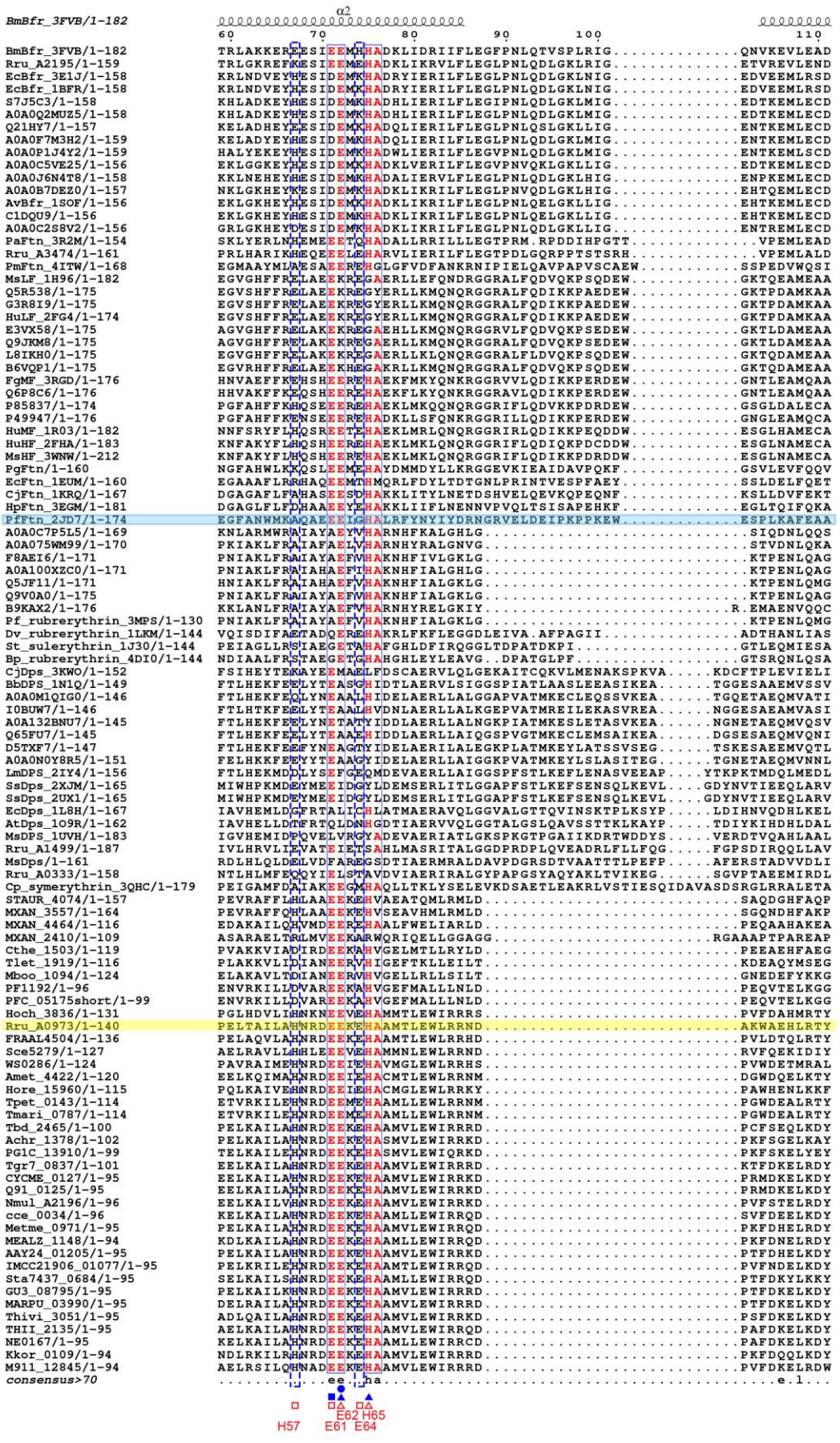
Rru-EncFtnSH-W38G	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMEELEAVDGYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13196.5	5.34	13980	1.059
Rru-EncFtnSH-E31A-W38G	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMAELEAVDGYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13138.47	5.45	13980	1.064
Rru-EncFtnSH-E31A-E34A-W38G	MAQSSNSTHEPLEVLKEETVNRHRAIVSVMAELAAVDGYD QRVDASTDPELTAILAHNRDEEKEHAAMTLEWLRRNDAKW AEHLRTYLFTEGPITAANSSSVDKLAAALEHHHHHH	116	13080.43	5.57	13980	1.069
Rru-Enc	MNDLMRDLAPISAKAWAEIETEARGTLTVTLAARKVVDFK GPLGWDASSVSLGRTEALAEEPKAAGSAAVVTVRKRAVQP LIELCVPFTLKRAELEAIARGASDADLDPVIEAARAIAIA EDRAVFHGFAAGGITGIGEASAEHALDLPADLADFPGVLV RALAVLRDRGVDGPYALVLGRTVYQQLMETTTPGGYPVLQ	280	29732.99	5.21	24075	0.81

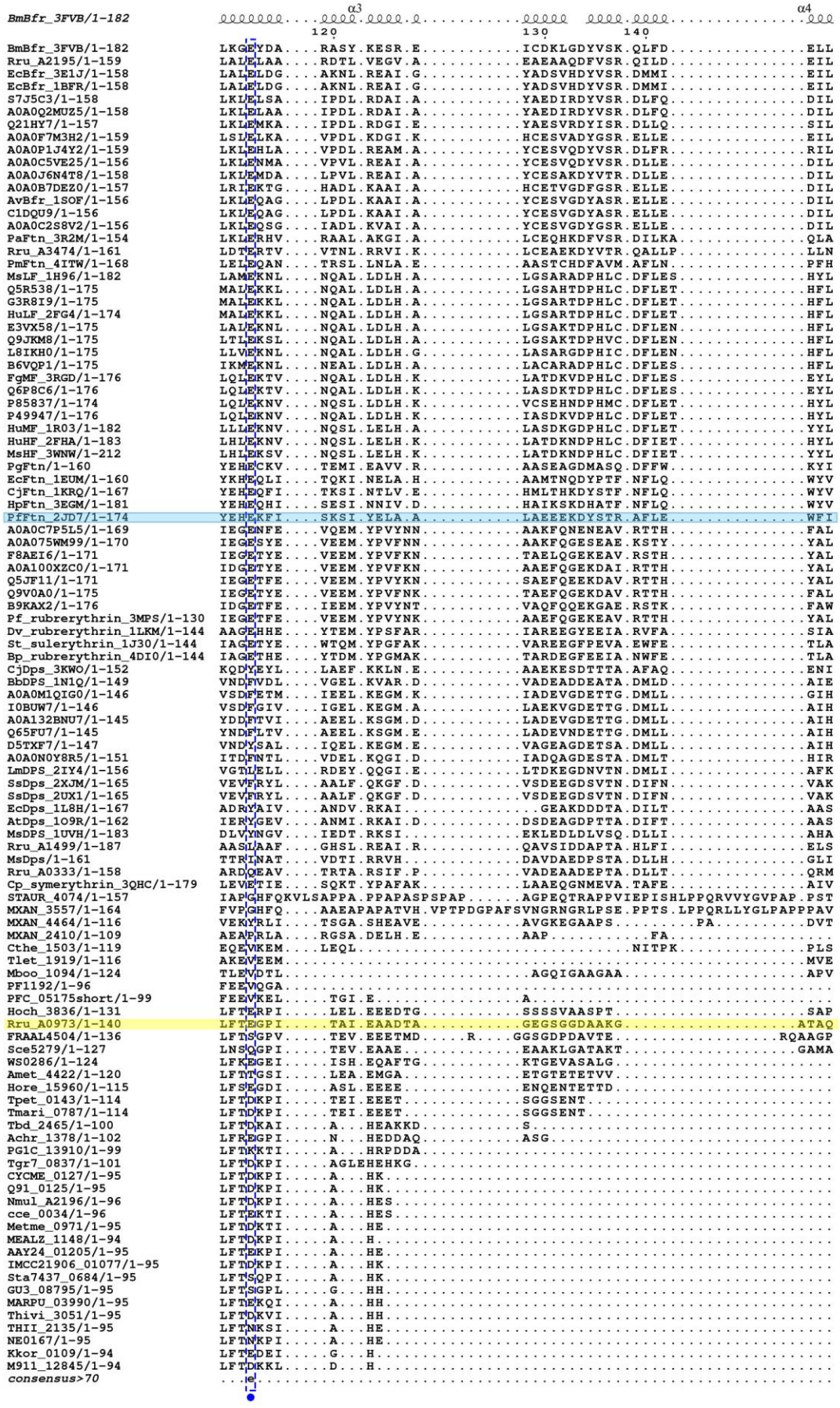
	HVRRLFEGPLIWAPGVDGAMLISQRGGDFELTVGRDFSIG YHDHDAQSVHLYLQESMTFRCLGPEAAVPLRGLSQAATKA					
Hoch-EncFtnS	MGSSEQLHEPAELLSEETKNMHRALVTLIEELEAVDWYQQ RADACSEPGLHDVLIHNKNEEVEHAMMTLEWIRRRSPVFD AHMRTYLFTERPILELEE	98	11572.03	4.76	13980	1.208
Pfc-EncFtnS	MGLSINPTLINRDKPYTKEELMEILRLAIIAELDAINLYE QMARYSEDENVRKILLDVAREEKAHVGEFMALLLNLDPEQ VTELKGGFEEVKELTGIEA	99	11304.01	4.49	4470	0.395

Table 7.2.1 Protein information.

Pfc-EncFtnS does not contain any Trp residues, which could result in more than 10% error in the computed extinction coefficient. The extinction coefficient was estimated assuming all pairs of Cys residues form cysteines.


BmBfr_3FVB/1-182
α1
1
10
20
30
40
50
BmBfr_3FVB/1-182
Rru_A2195/1-159
EcBfr_3E1J/1-158
EcBfr_1BFR/1-158
S7J5C3/1-158
A0A0Q2MUZ5/1-158
Q21HY7/1-157
A0A0F7M3H2/1-159
A0A0P1J4Y2/1-159
A0A0C5VE25/1-156
A0A0J6N4T8/1-158
A0A0B7DEZ0/1-157
AvBfr_1SOF/1-156
C1DQU9/1-156
A0A0C2S8V2/1-156
PaFtn_3R2M/1-154
Rru_A3474/1-161
PmFtn_4ITW/1-168
MsLF_1H96/1-182
Q5R538/1-175
G3R8I9/1-175
HuLF_2FG4/1-174
E3VX58/1-175
Q9JKM8/1-175
L8IKH0/1-175
B6VQP1/1-175
FgMF_3RGD/1-176
Q6P8C6/1-176
P85837/1-174
P49947/1-176
HuMF_1R03/1-182
HuHF_2FHA/1-183
MsHF_3WNW/1-212
PgFtn/1-160
EcFtn_1EUM/1-160
CjFtn_1KRQ/1-167
HpFtn_3EGM/1-181
PfFtn_2JD7/1-174
A0A0C7P5L5/1-169
A0A075WM99/1-170
F8AEI6/1-171
A0A100XZC0/1-171
Q5JF11/1-171
Q9V0A0/1-175
B9KAX2/1-176
Pf_rubrerythrin_3MPS/1-130
Dv_rubrerythrin_1LKM/1-144
St_sulerythrin_1J30/1-144
Bp_rubrerythrin_4DI0/1-144
CjDps_3KWO/1-152
BbDPS_1N1Q/1-149
A0A0M1QIG0/1-146
I0BUW7/1-146
A0A132BNU7/1-145
Q65FU7/1-145
D5TXF7/1-147
A0A0N0Y8R5/1-151
LmDPS_2IY4/1-156
SsDps_2XJM/1-165
SsDps_2UX1/1-165
EcDps_1L8H/1-167
AtDps_1O9R/1-162
MsDPS_1UVH/1-183
Rru_A1499/1-187
MsDps/1-161
Rru_A0333/1-158
Cp_symerythrin_3QHC/1-179
STAUR_4074/1-157
MXAN_3557/1-164
MXAN_4464/1-116
MXAN_2410/1-109
Cthe_1503/1-119
Tlet_1919/1-116
Mboo_1094/1-124
PF1192/1-96
PFC_05175short/1-99
Hoch_3836/1-131
Rru_A0973/1-140
FRAAL4504/1-136
Sce5279/1-127
WS0286/1-124
Amet_4422/1-120
Hore_15960/1-115
Tpet_0143/1-114
Tmari_0787/1-114
Tbd_2465/1-100
Achr_1378/1-102
PGlC_13910/1-99
Tgr7_0837/1-101
CYCME_0127/1-95
Q91_0125/1-95
Nmul_A2196/1-96
cce_0034/1-96
Metme_0971/1-95
MEALZ_1148/1-94
AAY24_01205/1-95
IMCC21906_01077/1-95
Sta7437_0684/1-95
GU3_08795/1-95
MARPU_03990/1-95
Thivi_3051/1-95
THII_2135/1-95
NE0167/1-95
Kkor_0109/1-94
M911_12845/1-94
consensus>70
E31
E32
E34
Y39


α2
BmBfr_3FVB/1-182
60
70
80
90
100
110
BmBfr_3FVB/1-182 TRLAKKERESIEEMHHADKLIDRIIFLEGFPNLQTVSPLRIG QNVKEVLEAD
Rru_A2195/1-159 TRLGKREFKESIEEMEHADKLIKRVLFLEGLPNLQDLGKLRIG ETVREVLEND
EcBfr_3E1J/1-158 KRLNDVEYHESIDEMKHADRYIERILFLEGLPNLQDLGKLNIG EDVEEMLRSD
EcBfr_1BFR/1-158 KRLNDVEYHESIDEMKHADRYIERILFLEGLPNLQDLGKLNIG EDVEEMLRSD
S7J5C3/1-158 KHLADKEYHESIDEMKHADHLIERILFLEGIPNLQDLGKLMIG EDTKEMLECD
A0A0Q2MUZ5/1-158 KHLADKEYHESIDEMKHADHLIERILFLEGLPNLQDLGKLMIG EDTKEMLECD
Q21HY7/1-157 KELADHEYEESIDEMKHADQLIERILFLEGLPNLQSLGKLLIG ENTKEMLECD
A0A0F7M3H2/1-159 KELADYEYHESIDEMKHADQLIERILFLEGLPNLQDLGKLMIG ENTKEMLECD
A0A0P1J4Y2/1-159 HALYEKEYHESIDEMKHADWLIERILFLEGVPNLQDLGKLMIG ENTKEMLSCD
A0A0C5VE25/1-156 EKLGGKEYHESIDEMKHADKLVERILFLEGVPNVQKLGKLLIG EDTKEMLECD
A0A0J6N4T8/1-158 KKLNEHEYHESIDEMKHADALIERVLFLEGLPNLQDLGKLHIG ENPKEMLECD
A0A0B7DEZ0/1-157 NKLGKHEYKESIDEMKHADKLIKRILFLEGLPNVQDLGKLHIG EHTQEMLECD
AvBfr_1SOF/1-156 EKLGKHEYHESIDEMKHADKLIKRILFLEGLPNLQELGKLLIG EHTKEMLECD
C1DQU9/1-156 EKLGKHEYHESIDEMKHADKLIKRILFLEGLPNLQELGKLLIG EHTKEMLECD
A0A0C2S8V2/1-156 GRLGKHEYDESIDEMKHADKLIKRILFLEGLPNLQDLGKLLIG ENTREMLECD
PaFtn_3R2M/1-154 SKLYERLNHEMEEETQHADALLRRILLLEGTPRM.RPDDIHPGTT VPEMLEAD
Rru_A3474/1-161 PRLHARIKHEQEEELEHARVLIERILFLEGTPDLGQRPPTSTSRH VPEMLALD
PmFtn_4ITW/1-168 EGMAAYMLAESAEEREHGLGFVDFANKRNIPIELQAVPAPVSCAEW SSPEDVWQSI
MsLF_1H96/1-182 EGVGHFFRELAEEKREGAERLLEFQNDRGGRALFQDVQKPSQDEW GKTQEAMEAA
Q5R538/1-175 EGVSHFFRELAEEKREGYERLLKMQNQRGGRALFQDIKKPAEDEW GKTPDAMKAA
G3R8I9/1-175 EGVSHFFRELAEEKREGYERLLKMQNQRGGRALFQDIKKPAEDEW GKTPDAMKAA
HuLF_2FG4/1-174 EGVSHFFRELAEEKREGYERLLKMQNQRGGRALFQDIKKPAEDEW GKTPDAMKAA
E3VX58/1-175 AGVGHFFRELAKEKREGAEHLLKMQNQRGGRVLFQDVQKPSEDEW GKTLDAMEAA
Q9JKM8/1-175 AGVGHFFRELAKEKREGAERLLKMQNQRGGRALFQDVQKPSEDEW GKTLDAMEAA
L8IKH0/1-175 EGVGHFFRELAKEKREGAERLLKLQNQRGGRALFLDVQKPSQDEW GKTQDAMEAA
B6VQP1/1-175 EGVRHFFRELAEEKHEGAKRLLKMQNQRGGRALFQDVQKPSQDEW GKTQDAMEAA
FgMF_3RGD/1-176 HNVAEFFKEHSHEEREHAEKFMKYQNKRGGRVVLQDIKKPERDEW GNTLEAMQAA
Q6P8C6/1-176 HHVAKFFKEQSHEERERHAEKFLKYQNKRGGRAVLQDIKKPERDEW GNTLEAMQAA
P85837/1-174 PGFAHFFKHQSEEEREHAEKLMKQQNQRGGRIFLQDVKKPDRDEW GSGLDALECA
P49947/1-176 PGFAHFFKENSEEEREHADKLLSFQNKRGGRILLQDIKKPERDEW GNGLEAMQCA
HuMF_1R03/1-182 NNFSRYFLHQSREETEHAEKLMRLQNQRGGRIRLQDIKKPEQDDW ESGLHAMECA
HuHF_2FHA/1-183 KNFAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIQKPDCDDW ESGLNAMECA
MsHF_3WNW/1-212 KNFAKYFLHQSHEEREHAEKLMKLQNQRGGRIFLQDIKKPDRDDW ESGLNAMECA
PgFtn/1-160 NGFAHWLKKQSLEEMEHAYDMMDYLLKRGGEVKIEAIDAVPQKF GSVLEVFQQV
EcFtn_1EUM/1-160 EGAAAFLRRHAQEEMTHMQRLFDYLTDTGNLPRINTVESPFAEY SSLDELFQET
CjFtn_1KRQ/1-167 DGAGAFLFAHASEESDHAKKLITYLNETDSHVELQEVKQPEQNF KSLLDVFEKT
HpFtn_3EGM/1-181 DGAGLFLFDHAAEEYEHAKKLIIFLNENNVPVQLTSISAPEHKF EGLTQIFQKA
PfFtn_2JD7/1-174 EGFANWMKAQAEEEIGHALRFYNYIYDRNGRVELDEIPKPPKEW ESPLKAFEAA
A0A0C7P5L5/1-169 KNLARMWRAIAYAEYVHARNHFKALGHLG SIQDNLQQS
A0A075WM99/1-170 PKIAKLFRAIAFAELVHARNHYRALGNVG STVDNLQKA
F8AEI6/1-171 PNIAKLFRAIAYAEFVHAKNHFIVLGKLG KTPENLQAG
A0A100XZC0/1-171 PNIAKLFRAIAHAEFIHAKNHFIALGHLG KTPENLQAG
Q5JF11/1-171 HNIAKLFRAIAHAEFVHAKNHFIALGKLG KTPENLQMG
Q9V0A0/1-175 PNIAKLFRAIAYAEFVHAKNHFIALGKLG KTPENLQAG
B9KAX2/1-176 KKLANLFRAIAYAEFVHARNHYRELGKIY R.EMAENVQQC
Pf_rubrerythrin_3MPS/1-130 PNIAKLFRAIAYAEFVHAKNHFIALGKLG KTPENLQMG
Dv_rubrerythrin_1LKM/1-144 VQISDIFAETADQERDHAKRLFKFLEGGDLEIVA.AFPAGII ADTHANLIAS
St_sulerythrin_1J30/1-144 PEIAGLLRSIAEGETAHAFGHLDFIRQGGLTDPATDKPI GTLEQMIESA
Bp_rubrerythrin_4DI0/1-144 NDIAALFRSTAEGETGHAHGHLEYLEAVG..DPATGLPF GTSRQNLQSA
CjDps_3KWO/1-152 FSIHEYTEKAYEEMAELFDSCAERVLQLGEKAITCQKVLMENAKSPKVA KDCFTPLEVIELI
BbDPS_1N1Q/1-149 FTLHEKFEELYTEASGHIDTLAERVLSIGGSPIATLAASLEEASIKEA TGGESAAEMVSSV
A0A0M1QIG0/1-146 FTLHEKFEQLYNEAALHIDELAERLLALGGAPVATMKECLEQSSVKEA TGQETAEQMVATI
I0BUW7/1-146 FTLHTKFEELYTEAALHVDNLAERLLALKGEPVATMREILEKASVREA NGGESAEAMVASI
A0A132BNU7/1-145 FTLHEKFEELYNETATYIDDLAERLLALNGKPIATMKESLETASVKEA NGNETAEQMVQSV
Q65FU7/1-145 FTLHEKFEELYTEAAEHIDDLAERLLAIQGSPVGTMKECLEMSAIKEA DGSESAEQMVQNI
D5TXF7/1-147 FTLHEKFEEFYNEAGTYIDELAERILALEGKPLATMKEYLATSSVSEG TSKESAEEMVQTL
A0A0N0Y8R5/1-151 FELHKKFEEYYTEAAGYIDELAERLLAIGGKPVATMKEYLSLASITEG TGNETAEQMVNNL
LmDPS_2IY4/1-156 FTLHEKMDDLYSEFGEQMDEVAERLLAIGGSPFSTLKEFLENASVEEAP YTKPKTMDQLMEDL
SsDps_2XJM/1-165 MIWHPKMDEYMEEIDGYLDEMSERLITLGGAPFSTLKEFSENSQLKEVL GDYNVTIEEQLARV
SsDps_2UX1/1-165 MIWHPKMDEYMEEIDGYLDEMSERLITLGGAPFSTLKEFSENSQLKEVL GDYNVTIEEQLARV
EcDps_1L8H/1-167 IAVHEMLDGFRTALICHLATMAERAVQLGGVALGTTQVINSKTPLKSYP LDIHNVQDHLKEL
AtDps_1O9R/1-162 IAVHELLDTFRTQLDNHGDTIAERVVQLGGTALGSLQAVSSTTKLKAYP TDIYKIHDHLDAL
MsDPS_1UVH/1-183 IGVHEMIDPQVELVRGYADEVAERIATLGKSPKGTPGAIIKDRTWDDYS VERDTVQAHLAAL
Rru_A1499/1-187 IVLHRVLIEVATEIETSAHLMASRITALGGDPRDPLQVEADRLFLLFQG FGPSDIRQQLLAV
MsDps/1-161 RDLHLQLDELVDFAREGSDTIAERMRALDAVPDGRSDTVAATTTLPEFP AFERSTADVVDLI
Rru_A0333/1-158 NTLHLMFEQQYIELSTAVDVIAERIRALGYPAPGSYAQYAKLTVIKEG SGVPTAEEMIRDL
Cp_symerythrin_3QHC/1-179 PEIGAMFDAIAKEEGMHAQLLTKLYSELEVKDSAETLEAKRLVSTIESQIDAVASDSRGLRRALETA
STAUR_4074/1-157 PEVRAFFLHLAAEEKEHVAEATQMLRMLD SAQDGHFAQP
MXAN_3557/1-164 PEVRAFFQHLAAEEKEHVSEAVHMLRMLD SGQNDHFAKP
MXAN_4464/1-116 EDAKAILQHVMDEERERHAALFWELIARLD PEQAAHAKEA
MXAN_2410/1-109 ASARAELTRLMVEEKARWQRIQELLGGAGG RGAAAPTPAREAP
Cthe_1503/1-119 PVAKKVIADIRDEEKAHVGELMTLLRYLD PEEAEHFAEG
Tlet_1919/1-116 PLAKKVLIDIANEERVHIGEFTKLLEILT KDEAQYMSEG
Mboo_1094/1-124 ELAKAVLTDIANEERVHVGELLRLLSILT GNEDEFYKKG
PF1192/1-96 ENVRKILLDVAREEKAHVGEFMALLLNLD PEQVTELKGG
PFC_05175short/1-99 ENVRKILLDVAREEKAHVGEFMALLLNLD PEQVTELKGG
Hoch_3836/1-131 PGLHDVLIHHNKNEEVEHAMMTLEWIRRRS PVFDAHMRTY
Rru_A0973/1-140 PELTAILAHNRDEEKEHAAMTLEWLRRND AKWAEHLRTY
FRAAL4504/1-136 PELAQVLAHNRDEEKEHAAMTLEWLRRHD PVLDTQLRTY
Sce5279/1-127 AELRAVLLHHLEEEVEHAMMNLEWLRRND RVFQEKIDIY
WS0286/1-124 PAVRAIMEHNRDEEIEHAVMGLEWLRRHS PVWDETMRAL
Amet_4422/1-120 EELKQIMAHNRDEEIEHACMTLEWLRRNM DGWDQELKTY
Hore_15960/1-115 PQLKAIVEHNRDEEIEHACMGLEWLRRKY PAWHENLKKF
Tpet_0143/1-114 ETVRKILEHNRDEEMEHAAMLLEWLRRNM PGWDEALRTY
Tmari_0787/1-114 ETVRKILEHNRDEEMEHAAMLLEWLRRNM PGWDEALRTY
Tbd_2465/1-100 PELKAILAHNRDEEKEHAAMVLEWIRRRD PCFSEQLKDY
Achr_1378/1-102 PELKAILAHNRDEEKEHASMVLEWIRRKD PKFSGELKAY
PG1C_13910/1-99 TELKAILEHNRDEEKEHASMVLEWIQRKD PKFSKELYEY
Tgr7_0837/1-101 EELKAILAHNRDEEKEHAAMVLEWIRRRD KTFDKELRDY
CYCME_0127/1-95 EELKAILAHNRDEEKEHAAMVLEWIRRKD PRMDKELKDY
Q91_0125/1-95 EELKAILAHNRDEEKEHAAMVLEWIRRKD PRMDKELKDY
Nmul_A2196/1-96 EELKAILVHNRDEEKEHAAMVLEWIRRKD PVFSTELRDY
cce_0034/1-96 KELKAILAHNRDEEKEHAAMLLEWIRRQD SVFDEELKDY
Metme_0971/1-95 PELKAILAHNRDEEKEHAAMVLEWIRRQD PKFDHELRDY
MEALZ_1148/1-94 KDLKAILAHNRDEEKEHAAMLLEWIRRQD PKFDNELRDY
AAY24_01205/1-95 PELKAILAHNRDEEKEHAAMVLEWIRRKD PTFDHELKDY
IMCC21906_01077/1-95 PELKRILEHNRDEEKEHAAMTLEWIRRQD PVFNDELKDY
Sta7437_0684/1-95 SELKAILSHNRDEEKEHAAMLLEWIRRQD PTFDKYLKKY
GU3_08795/1-95 PELKAILKHNRDEEKEHAMVLEWIRRRD PTFDKELKDY
MARPU_03990/1-95 DELRAILAHNRDEEKEHAAMVLEWIRRRD PTFDKELKDY
Thivi_3051/1-95 ADLQAILAHNRDEEKEHAAMVLEWIRRKD PSFDKELKDY
THII_2135/1-95 AELKAILAHNRDEEKEHAAMVLEWIRRQD PAFDKELKDY
NE0167/1-95 KELKAILAHNRDEEKEHAAMLLEWIRRCD PAFDKELKDY
Kkor_0109/1-94 NDLRKILRHNRDEEKEHAAMVLEWIRRRD PKFDQQLKDY
M911_12845/1-94 AELRSILQHNADEEKEHAAMVLEWIRRRD PVFDKELRDW
consensus>70 ....ee..ha....e.l...
H57
E61
E62
E64
H65


BmBfr_3FVB/1-182
α3
α4
120
130
140
BmBfr_3FVB/1-182
Rru_A2195/1-159
EcBfr_3E1J/1-158
EcBfr_1BFR/1-158
S7J5C3/1-158
A0A0Q2MUZ5/1-158
Q21HY7/1-157
A0A0F7M3H2/1-159
A0A0P1J4Y2/1-159
A0A0C5VE25/1-156
A0A0J6N4T8/1-158
A0A0B7DEZ0/1-157
AvBfr_1SOF/1-156
C1DQU9/1-156
A0A0C2S8V2/1-156
PaFtn_3R2M/1-154
Rru_A3474/1-161
PmFtn_4ITW/1-168
MsLF_1H96/1-182
Q5R538/1-175
G3R8I9/1-175
HuLF_2FG4/1-174
E3VX58/1-175
Q9JKM8/1-175
L8IKH0/1-175
B6VQP1/1-175
FgMF_3RGD/1-176
Q6P8C6/1-176
P85837/1-174
P49947/1-176
HuMF_1R03/1-182
HuHF_2FHA/1-183
MsHF_3WNW/1-212
PgFtn/1-160
EcFtn_1EUM/1-160
CjFtn_1KRQ/1-167
HpFtn_3EGM/1-181
PfFtn_2JD7/1-174
A0A0C7P5L5/1-169
A0A075WM99/1-170
F8AEI6/1-171
A0A100XZC0/1-171
Q5JF11/1-171
Q9V0A0/1-175
B9KAX2/1-176
Pf_rubrerythrin_3MPS/1-130
Dv_rubrerythrin_1LKM/1-144
St_sulerythrin_1J30/1-144
Bp_rubrerythrin_4DI0/1-144
CjDps_3KWO/1-152
BbDPS_1N1Q/1-149
A0A0M1QIG0/1-146
I0BUW7/1-146
A0A132BNU7/1-145
Q65FU7/1-145
D5TXF7/1-147
A0A0N0Y8R5/1-151
LmDPS_2IY4/1-156
SsDps_2XJM/1-165
SsDps_2UX1/1-165
EcDps_1L8H/1-167
AtDps_1O9R/1-162
MsDPS_1UVH/1-183
Rru_A1499/1-187
MsDps/1-161
Rru_A0333/1-158
Cp_symerythrin_3QHC/1-179
STAUR_4074/1-157
MXAN_3557/1-164
MXAN_4464/1-116
MXAN_2410/1-109
Cthe_1503/1-119
Tlet_1919/1-116
Mboo_1094/1-124
PF1192/1-96
PFC_05175short/1-99
Hoch_3836/1-131
Rru_A0973/1-140
FRAAL4504/1-136
Sce5279/1-127
WS0286/1-124
Amet_4422/1-120
Hore_15960/1-115
Tpet_0143/1-114
Tmari_0787/1-114
Tbd_2465/1-100
Achr_1378/1-102
PG1C_13910/1-99
Tgr7_0837/1-101
CYCME_0127/1-95
Q91_0125/1-95
Nmul_A2196/1-96
cce_0034/1-96
Metme_0971/1-95
MEALZ_1148/1-94
AAY24_01205/1-95
IMCC21906_01077/1-95
Sta7437_0684/1-95
GU3_08795/1-95
MARPU_03990/1-95
Thivi_3051/1-95
THII_2135/1-95
NE0167/1-95
Kkor_0109/1-94
M911_12845/1-94
consensus>70

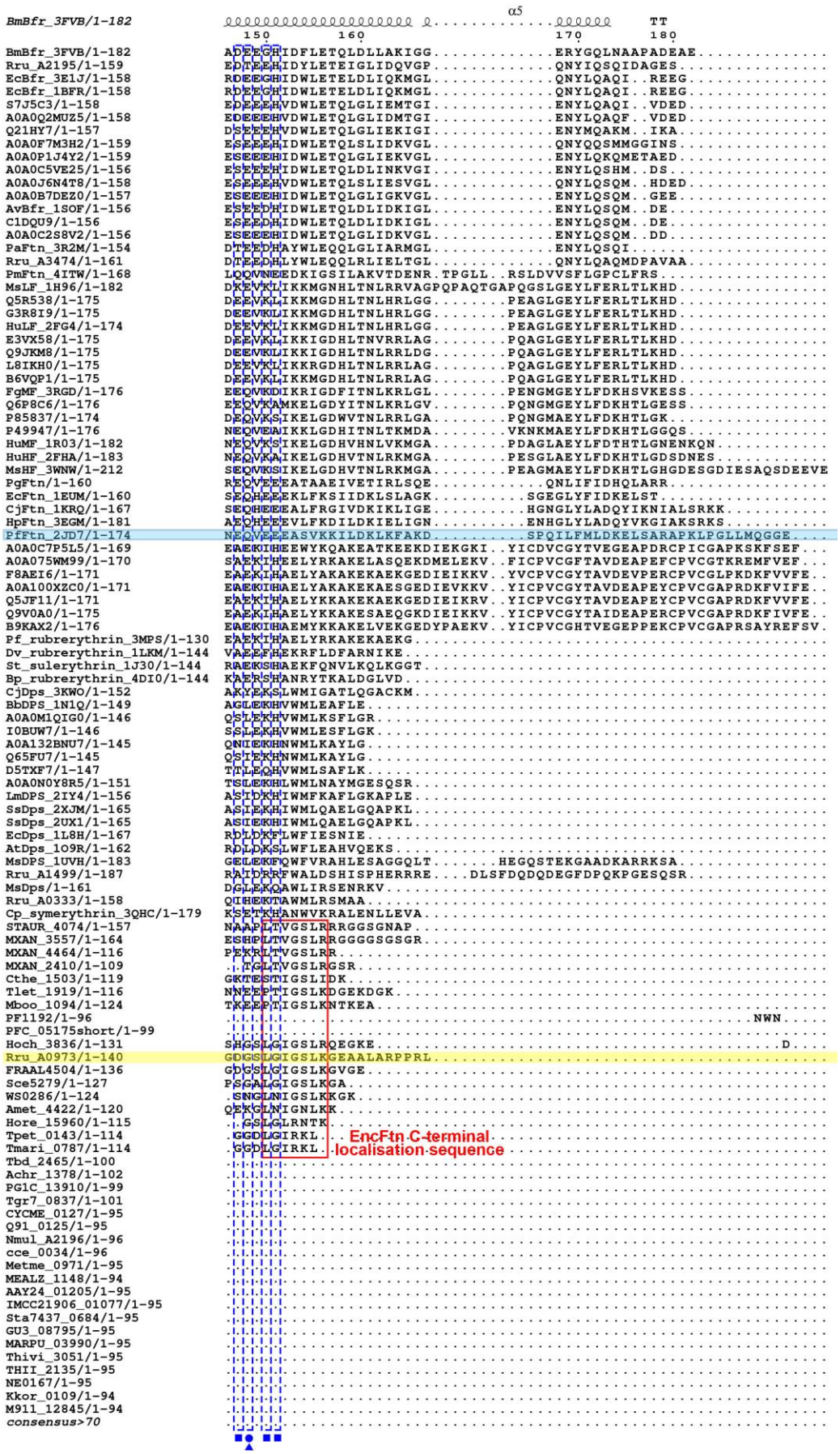

α5
BmBfr_3FVB/1-182
150
160
170
180
EncFtn C-terminal
localisation sequence

Figure 7.2.1 Multiple protein sequence alignment of ferritin family proteins.

Protein names were adapted from either UniprotKB (UniProt Consortium 2014), KEGG (Ogata et al. 1999) database or common name with PDB entry code (Berman et al. 2000). Sequences were sorted in an order corresponding to the clades in phylogenetic tree (**Figure 1.4.9**). The alignment was performed progressively by EBI alignment tools as described above and edited by Espript 3.0 web server (Robert & Gouet 2014). The secondary structure of *Brucella melitensis* BFR (BmBfr, PDB ID: 3FVB) is shown on top of the alignment. Rru-EncFtn (Rru_A0973), as a representative of EncFtn proteins, is highlighted in yellow. The metal binding sites of *P. furiosus* bacterial ferritin (PfFtn_2JD7) (highlighted in blue) consist of Fe_A site (E17, E50, H53 and Q127, highlighted as solid blue triangles), Fe_B site (E50, E94, E130, Q127, labelled with solid blue circles) and Fe_C site (E49, E126, E129 and E130, solid blue squares). The FOC of Rru_A0973 was labelled with empty red triangles including E32, E62, H65 and Y39; and the metal entry site was marked with empty red circles including E31 and E34.The metal exit site was marked with empty red squares including H57, E61 and E64. The C-terminal localisation sequences shared by the encapsulin associated ferritins (EncFtn) were highlighted within the red rectangle.

FOC-dimer interface summary	Rru-EncFtnSH				Pfc-EncFtnS				Hoch-EncFtnS			
	Chain D		Chain E		Chain a		Chain c		Chain E		Chain I	
Number of atoms												
interface	138	18.50%	142	19.20%	179	23.00%	178	22.90%	132	17.20%	127	16.70%
surface	558	74.80%	555	74.90%	567	72.90%	574	73.80%	564	73.40%	548	72.20%
total	746	100.00%	741	100.00%	778	100.00%	778	100.00%	768	100.00%	759	100.00%
Number of residues												
interface	33	36.30%	31	34.40%	44	45.40%	44	45.40%	31	33.70%	31	34.10%
surface	91	100.00%	90	100.00%	97	100.00%	97	100.00%	91	98.90%	89	97.80%
total	91	100.00%	90	100.00%	97	100.00%	97	100.00%	92	100.00%	91	100.00%
Solvent-accessible area, Å												
interface	1240.7	17.00%	1243.3	16.90%	1796.1	22.50%	1797.1	21.90%	1208.8	16.60%	1224.9	16.90%
total	7304.3	100.00%	7352.3	100.00%	7977.7	100.00%	8197.5	100.00%	7302.1	100.00%	7266.3	100.00%
Solvation energy, kcal/mol												
isolated structure	-56.4	100.00%	-53.9	100.00%	-65.3	100.00%	-64.7	100.00%	-66.5	100.00%	-64.6	100.00%
gain on complex	-5.8	10.30%	-5.6	10.40%	-12.7	19.50%	-12.7	19.70%	-5	7.60%	-3.8	5.90%
average gain	-6.1	10.70%	-6.6	12.30%	-10.6	16.20%	-10.6	16.40%	-6	8.90%	-6.4	9.90%
P-value	0.554		0.623		0.34		0.351		0.601		0.751	
Non-FOC-dimer interface summary	**Rru-EncFtnSH**				**Pfc-EncFtnS**				**Hoch-EncFtnS**			
	Chain A		Chain B		Chain O		Chain Z		Chain E		Chain G	
Number of atoms												
interface	268	35.90%	262	35.10%	195	25.10%	194	24.80%	241	31.40%	240	31.20%
surface	556	74.50%	549	73.60%	561	72.10%	573	73.20%	552	71.90%	560	72.90%
total	746	100.00%	746	100.00%	778	100.00%	783	100.00%	768	100.00%	768	100.00%
Number of residues												
interface	55	60.40%	55	60.40%	46	47.40%	46	46.90%	54	58.70%	53	57.60%
surface	91	100.00%	91	100.00%	97	100.00%	98	100.00%	91	98.90%	89	96.70%
total	91	100.00%	91	100.00%	97	100.00%	98	100.00%	92	100.00%	92	100.00%
Solvent-accessible area, Å												
interface	2457.5	33.80%	2468.3	33.80%	1816	22.80%	1806.7	22.50%	2326.9	31.90%	2342.4	32.30%
total	7281	100.00%	7307.7	100.00%	7954.7	100.00%	8026.8	100.00%	7296.6	100.00%	7245.3	100.00%
Solvation energy, kcal/mol												
isolated structure	-55.3	100.00%	-56.2	100.00%	-62.1	100.00%	-67	100.00%	-66.7	100.00%	-65.3	100.00%
gain on complex	-15.1	27.30%	-15.6	27.70%	-7	11.30%	-6.5	9.70%	-17.2	25.80%	-16.5	25.30%
average gain	-12.3	22.30%	-11.8	21.00%	-12.7	20.50%	-11.2	16.70%	-11	16.60%	-11.6	17.70%
P-value	0.314		0.254		0.875		0.831		0.124		0.173	

Table 7.2.2 Interaction summary of dimer interfaces of EncFtn homologues analysed by PDBePISA server.

Hydrogen bonds in FOC-dimer interface

Rru-EncFtnSH				Pfc-EncFtnS			
No.	Chain D	Dist. [Å]	Chain E	No.	Chain a	Dist. [Å]	Chain c
1	D:TYR 39[HH]	1.82	E:GLU 32[OE2]	1	a:LEU 3[H]	2.04	c:ASP 34[OD2]
2	D:THR 47[H]	2.38	E:ASN 21[OD1]	2	a:SER 4[H]	2.09	c:ASP 34[OD1]
3	D:HIS 65[ND1]	3.66	E:GLU 62[OE2]	3	a:ILE 5[H]	2.19	c:GLN 41[OE1]
4	D:THR 47[OG1]	2.02	E:ASN 21[HD21]	4	a:TYR 16[H]	1.94	c:TYR 45[O]
5	D:GLU 32[OE2]	1.76	E:TYR 39[HH]	5	a:TYR 39[HH]	1.90	c:GLU 32[OE2]
6	D:ASN 21[OD1]	3.88	E:SER 46[OG]	6	a:GLN 41[HE22]	1.96	c:ILE 5[O]
7	D:ASN 21[OD1]	2.34	E:THR 47[H]	7	a:TYR 45[HH]	2.14	c:ARG 12[O]
8	D:GLU 62[OE2]	3.49	E:HIS 65[ND1]	8	a:ASP 34[OD2]	2.04	c:LEU 3[H]
Hoch-EncFtnS				9	a:ASP 34[OD1]	2.19	c:SER 4[H]
No.	Chain E	Dist. [Å]	Chain I	10	a:GLN 41[OE1]	2.22	c:ILE 5[H]
1	E:TYR 38[HH]	2.16	I:GLU 31[OE2]	11	a:TYR 45[O]	1.91	c:TYR 16[H]
2	E:ARG 41[HH12]	1.96	I:GLU 30[OE1]	12	a:GLU 32[OE2]	1.86	c:TYR 39[HH]
3	E:ARG 41[HH22]	1.69	I:GLU 30[OE2]	13	a:ILE 5[O]	1.90	c:GLN 41[HE22]
4	E:GLU 30[OE1]	1.78	I:ARG 41[HH22]	14	a:ARG 12[O]	2.06	c:TYR 45[HH]
5	E:GLU 47[OE2]	2.20	I:ARG 75[HH21]				

Salt bridges in FOC-dimer interface

Rru-EncFtnSH				Hoch-EncFtnS			
No.	Chain D	Dist. [Å]	Chain E	No.	Chain E	Dist. [Å]	Chain I
1	D:HIS 65[ND1]	3.85	E:GLU 62[OE1]	1	E:ARG 41[NH1]	2.81	I:GLU 30[OE1]
2	D:HIS 65[ND1]	3.66	E:GLU 62[OE2]	2	E:ARG 41[NH1]	2.88	I:GLU 30[OE2]
3	D:HIS 65[NE2]	2.76	E:GLU 61[OE1]	3	E:ARG 41[NH2]	3.85	I:GLU 30[OE1]
4	D:GLU 62[OE2]	3.49	E:HIS 65[ND1]	4	E:ARG 41[NH2]	2.45	I:GLU 30[OE2]
5	D:GLU 62[OE1]	3.71	E:HIS 65[ND1]	5	E:HIS 64[ND1]	3.77	I:GLU 61[OE2]
6	D:GLU 61[OE1]	2.50	E:HIS 65[NE2]	6	E:HIS 64[NE2]	2.77	I:GLU 60[OE1]
				7	E:ARG 75[NE]	3.98	I:GLU 47[OE1]
				8	E:GLU 30[OE1]	2.51	I:ARG 41[NH2]
Pfc-EncFtnS				9	E:GLU 30[OE1]	3.92	I:ARG 41[NH1]
No.	Chain a	Dist. [Å]	Chain c	10	E:GLU 30[OE2]	3.62	I:ARG 41[NH2]
1	a:HIS 65[ND1]	4.00	c:GLU 62[OE1]	11	E:GLU 30[OE2]	3.65	I:ARG 41[NH1]
2	a:HIS 65[ND1]	3.36	c:GLU 62[OE2]	12	E:GLU 47[OE1]	3.71	I:ARG 75[NH2]
3	a:HIS 65[NE2]	3.00	c:GLU 61[OE1]	13	E:GLU 47[OE2]	3.96	I:ARG 75[NE]
4	a:HIS 65[NE2]	3.53	c:GLU 61[OE2]	14	E:GLU 47[OE2]	2.98	I:ARG 75[NH2]
5	a:GLU 62[OE2]	3.34	c:HIS 65[ND1]	15	E:GLU 60[OE1]	2.99	I:HIS 64[NE2]

Table 7.2.3 Hydrogen bonds and salt bridges of FOC dimer interface of EncFtn homologues analysed by PDBePISA.

Salt bridges in Non-FOC-dimer interface							
Rru-EncFtnSH				**Pfc-EncFtnS**			
No.	Chain A	Dist. [Å]	Chain B	No.	Chain O	Dist. [Å]	Chain Z
1	A:HIS 23[ND1]	3.46	B:ASP 37[OD1]	1	O:ARG 26[NH1]	3.87	Z:GLU 40[OE1]
2	A:HIS 23[ND1]	2.59	B:ASP 37[OD2]	2	O:ARG 26[NH1]	2.88	Z:GLU 40[OE2]
3	A:ARG 22[NH2]	3.77	B:ASP 40[OD1]	3	O:ARG 26[NH2]	3.16	Z:GLU 40[OE1]
4	A:ARG 22[NH2]	2.66	B:ASP 40[OD2]	4	O:ARG 26[NH2]	3.57	Z:GLU 40[OE2]
5	A:ARG 22[NE]	3.26	B:ASP 40[OD2]	5	O:ARG 44[NH1]	2.86	Z:GLU 19[OE1]
6	A:LYS 63[NZ]	2.99	B:ASP 77[OD1]	6	O:ARG 44[NH1]	3.32	Z:GLU 23[OE2]
7	A:LYS 63[NZ]	3.24	B:ASP 77[OD2]	7	O:ARG 44[NH2]	3.05	Z:GLU 23[OE1]
8	A:ASP 37[OD1]	3.55	B:HIS 23[ND1]	8	O:ARG 44[NH2]	3.26	Z:GLU 23[OE2]
9	A:ASP 37[OD2]	2.59	B:HIS 23[ND1]	9	O:ARG 60[NH1]	3.34	Z:GLU 83[OE2]
10	A:ASP 40[OD1]	3.88	B:ARG 22[NH2]	10	O:ARG 60[NH1]	3.85	Z:GLU 83[OE1]
11	A:ASP 40[OD2]	3.41	B:ARG 22[NE]	11	O:LYS 63[NZ]	3.01	Z:ASP 77[OD1]
12	A:ASP 40[OD2]	2.82	B:ARG 22[NH2]	12	O:LYS 63[NZ]	2.97	Z:ASP 77[OD2]
13	A:ASP 44[OD1]	3.40	B:LYS 16[NZ]	13	O:LYS 63[NZ]	3.02	Z:GLU 79[OE1]
14	A:ASP 77[OD1]	2.91	B:LYS 63[NZ]	14	O:GLU 40[OE1]	3.91	Z:ARG 26[NH1]
15	A:ASP 77[OD2]	3.14	B:LYS 63[NZ]	15	O:GLU 40[OE2]	2.90	Z:ARG 26[NH1]
				16	O:GLU 40[OE1]	3.16	Z:ARG 26[NH2]
				18	O:GLU 20[OE2]	2.70	Z:ARG 44[NH1]
Hoch-EncFtnS				19	O:GLU 20[OE2]	3.45	Z:ARG 44[NH2]
No.	Chain E	Dist. [Å]	Chain G	20	O:GLU 83[OE1]	3.85	Z:ARG 60[NH1]
1	E:HIS 22[ND1]	3.35	G:ASP 36[OD1]	21	O:GLU 83[OE2]	3.68	Z:ARG 60[NH1]
2	E:HIS 22[ND1]	2.46	G:ASP 36[OD2]	22	O:ASP 77[OD1]	3.24	Z:LYS 63[NZ]
3	E:ASP 36[OD1]	3.44	G:HIS 22[ND1]	23	O:ASP 77[OD2]	3.20	Z:LYS 63[NZ]
4	E:ASP 36[OD2]	2.55	G:HIS 22[ND1]	24	O:GLU 79[OE1]	2.91	Z:LYS 63[NZ]

Table 7.2.4 Salt bridges in Non-FOC dimer interface of EncFtn homologues analysed by PDBePISA.

Hydrogen bonds in Non-FOC-dimer interface											
Rru-EncFtnSH				**Pfc-EncFtnS**				**Hoch-EncFtnS**			
No.	Chain A	Dist. [Å]	Chain B	No.	Chain O	Dist. [Å]	Chain Z	No.	Chain E	Dist. [Å]	Chain G
1	A:ARG 22[HH21]	1.84	B:ASP 40[OD2]	1	O:ARG 26[HH12]	2.10	Z:GLU 40[OE2]	1	E:HIS 8[H]	2.02	G:GLN 40[OE1]
2	A:HIS 9[H]	1.99	B:GLN 41[OE1]	2	O:ARG 26[HH21]	2.10	Z:ASN 37[OD1]	2	E:SER 15[H]	2.37	G:ASP 43[OD2]
3	A:HIS 9[ND1]	3.74	B:GLN 41[OE1]	3	O:ARG 26[HH22]	2.32	Z:GLU 40[OE1]	3	E:TYR 86[HH]	1.84	G:GLU 63[OE2]
4	A:LYS 16[H]	2.28	B:ASP 44[OD2]	4	O:ASN 37[HD21]	2.17	Z:GLU 23[O]	4	E:PHE 88[H]	2.21	G:GLU 70[OE1]
5	A:TYR 87[HH]	1.71	B:GLU 64[OE2]	5	O:GLN 41[HE21]	2.41	Z:GLU 23[OE1]	5	E:THR 89[H]	2.36	G:GLU 70[OE2]
6	A:PHE 89[H]	2.26	B:GLU 71[OE1]	6	O:ARG 44[HH11]	2.28	Z:GLU 19[OE1]	6	E:ARG 74[HH21]	1.89	G:THR 89[OG1]
7	A:THR 90[H]	2.44	B:GLU 71[OE2]	7	O:ARG 44[HH22]	2.27	Z:GLU 23[OE1]	7	E:ARG 74[HH12]	2.27	G:ARG 91[O]
8	A:LYS 63[HZ3]	2.11	B:ASP 77[OD1]	8	O:ARG 60[HH11]	2.49	Z:GLU 83[OE2]	8	E:ARG 74[HH22]	2.05	G:ARG 91[O]
9	A:ARG 75[HH21]	2.01	B:THR 90[OG1]	9	O:LYS 63[HZ3]	2.16	Z:ASP 77[OD1]	9	E:GLN 40[OE1]	1.93	G:HIS 8[H]
10	A:ARG 75[HH12]	2.10	B:GLY 92[O]	10	O:LYS 63[HZ2]	2.16	Z:GLU 79[OE1]	10	E:ASP 43[OD2]	2.49	G:SER 15[H]
11	A:ARG 75[HH22]	2.38	B:GLY 92[O]	11	O:GLU 40[OE2]	2.14	Z:ARG 26[HH12]	11	E:ASP 43[OD2]	3.59	G:SER 15[OG]
12	A:ASP 40[OD2]	2.00	B:ARG 22[HH21]	12	O:ASN 37[OD1]	2.04	Z:ARG 26[HH21]	12	E:GLU 63[OE2]	1.70	G:TYR 86[HH]
13	A:GLN 41[OE1]	1.87	B:HIS 9[H]	13	O:GLU 40[OE1]	2.31	Z:ARG 26[HH22]	13	E:GLU 70[OE1]	2.21	G:PHE 88[H]
14	A:ASP 44[OD2]	2.09	B:LYS 16[H]	14	O:GLU 23[O]	2.22	Z:ASN 37[HD21]	14	E:GLU 70[OE2]	2.28	G:THR 89[H]
15	A:GLU 64[OE2]	1.84	B:TYR 87[HH]	15	O:GLU 23[OE1]	2.11	Z:GLN 41[HE21]	15	E:THR 89[OG1]	2.05	G:ARG 74[HH21]
16	A:GLU 71[OE1]	2.29	B:PHE 89[H]	16	O:GLU 20[OE2]	1.98	Z:ARG 44[HH12]	16	E:ARG 91[O]	2.32	G:ARG 74[HH12]
17	A:GLU 71[OE2]	2.38	B:THR 90[H]	17	O:ASP 77[OD1]	2.39	Z:LYS 63[HZ3]	17	E:ARG 91[O]	2.20	G:ARG 74[HH22]
18	A:ASP 77[OD1]	2.04	B:LYS 63[HZ3]	18	O:ASP 77[OD2]	2.43	Z:LYS 63[HZ3]				
19	A:THR 90[OG1]	2.17	B:ARG 75[HH21]	19	O:GLU 79[OE1]	2.03	Z:LYS 63[HZ2]				
20	A:GLY 92[O]	2.33	B:ARG 75[HH22]								
21	A:GLY 92[O]	2.04	B:ARG 75[HH12]								

Table 7.2.5 Hydrogen bonds in non-FOC dimer interface of EncFtn homologues analysed by PDBePISA.

CheckMyMetal reports

1. Rru-EncFtnSH

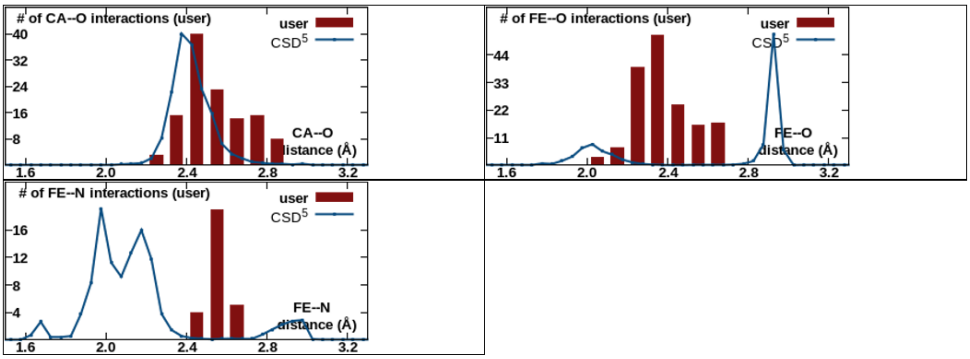
Warning: *Valence* and *nVECSUM* parameters should be interpreted with great care due to the presence of multi-nuclear metal clusters around e:1 e:5 e:7 e:8 e:11 e:12 e:13 e:15 e:17 e:18 e:19 e:21 e:22 e:23 e:25 e:28 e:29 i:1 i:4

ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
e:1	FE	Fe	1	23.3 (25.6)	O_6N_1	**1.1**	0.092	*Trigonal Bipyramidal*	11.4°	0	2	
e:2	FE	Fe	1	21.6 (24.4)	O_6N_1	*1.2*	0.044	*Trigonal Bipyramidal*	10.9°	0	2	
e:5	FE	Fe	1	*20 (24.2)*	O_5N_1	**1.1**	*0.15*	*Trigonal Bipyramidal*	12.3°	0	1	
e:6	FE	Fe	1	*20.8 (25.5)*	O_6N_1	*1.2*	0.052	*Trigonal Bipyramidal*	12.1°	0	2	
e:7	FE	Fe	1	*21.8 (25.6)*	O_5N_1	**1.1**	*0.11*	*Trigonal Bipyramidal*	12°	0	1	
e:8	FE	Fe	1	21.2 (24.4)	O_5N_1	**1.1**	*0.14*	*Trigonal Bipyramidal*	12.8°	0	1	
e:9	FE	Fe	1	21.8 (25.3)	O_6N_1	*1.2*	0.052	*Trigonal Bipyramidal*	11.9°	0	2	
e:10	FE	Fe	1	21.6 (24.4)	O_6N_1	*1.2*	0.039	*Trigonal Bipyramidal*	12.6°	0	2	
e:11	FE	Fe	1	*18.9 (24.7)*	O_5N_1	**1.1**	*0.15*	*Trigonal Bipyramidal*	12.6°	0	1	
e:12	FE	Fe	1	*20.8 (25.6)*	O_5N_1	**0.9**	0.091	*Trigonal Bipyramidal*	13.4°	0	1	
e:13	FE	Fe	1	*22.4 (26.3)*	O_5N_1	**1.1**	*0.17*	*Trigonal Bipyramidal*	*13.8°*	0	1	
e:14	FE	Fe	1	24.6 (27.3)	O_6N_1	*1.3*	0.057	*Trigonal Bipyramidal*	10.8°	0	2	
e:15	FE	Fe	1	21.7 (24.4)	O_5N_1	**1.1**	*0.19*	*Trigonal Bipyramidal*	10.7°	0	1	
e:16	FE	Fe	1	*20.6 (25)*	O_6N_1	*1.3*	0.1	*Trigonal Bipyramidal*	10.7°	0	2	
e:17	FE	Fe	1	23.7 (26.7)	O_5N_1	**1.1**	0.1	*Trigonal Bipyramidal*	12.3°	0	1	
e:18	FE	Fe	1	23.5 (25.7)	O_5N_1	**1**	*0.12*	*Trigonal Bipyramidal*	12.4°	0	1	
e:19	FE	Fe	1	23.8 (26)	O_5N_1	**1**	*0.21*	*Trigonal Bipyramidal*	12.1°	0	1	
e:20	FE	Fe	1	*24.8 (29.3)*	O_6N_1	*1.2*	*0.11*	*Trigonal Bipyramidal*	11.5°	0	2	
e:21	FE	Fe	1	23.2 (27)	O_5N_1	**1**	*0.14*	*Trigonal Bipyramidal*	11.3°	0	1	
e:22	FE	Fe	1	26.8 (28.7)	O_5N_1	**1.1**	*0.15*	*Trigonal Bipyramidal*	12.6°	0	1	
e:23	FE	Fe	1	26.8 (30.2)	O_5N_1	**1.1**	0.065	*Trigonal Bipyramidal*	11.2°	0	1	
e:24	FE	Fe	1	*23.1 (27.8)*	O_6N_1	*1.2*	0.061	*Trigonal Bipyramidal*	10.1°	0	2	
e:25	FE	Fe	1	27.6 (30)	O_4	**0.9**	*0.19*	**Trigonal Planar**	9.1°	0	1	
e:26	FE	Fe	1	*25.8 (31.7)*	O_5N_1	*1.4*	*0.15*	*Trigonal Bipyramidal*	10.2°	0	1	
e:27	FE	Fe	1	*23.8 (31.6)*	O_6N_1	*1.5*	*0.22*	*Trigonal Bipyramidal*	7.5°	0	2	
e:28	FE	Fe	1	28.1 (30.3)	O_5N_1	**1.1**	0.073	*Trigonal Bipyramidal*	5.8°	*20%*	2	
e:29	FE	Fe	1	36.4 (36.1)	O_4	**1.1**	0.08	**Trigonal Planar**	10.1°	0	1	
e:32	FE	Fe	1	34 (33.7)	O_5N_1	*1.4*	*0.15*	*Trigonal Bipyramidal*	12.8°	0	1	
e:33	FE	Fe	1	35.8	O_5N_1	*1.4*	0.061	*Trigonal*	13.4°	0	1	

e:33	FE	Fe	1	(38.9)	O_5N_1	1.4	0.081	Bipyramidal	13.4	0	1	
e:45	FE	Fe	1	34.7 (36.6)	O_5N_1	1.4	0.17	Trigonal Bipyramidal	11.5°	0	1	
i:1	CA	Ca	1	84.6 (49.5)	O_3	0.5	0.7	Octahedral	14.3°	50%	0	
i:2	CA	Ca	1	75.5 (45)	O_1	0.15	1	Poorly Coordinated	N/A	N/A	0	
i:3	CA	Ca	1	85.1 (55.4)	O_1	0.13	1	Poorly Coordinated	N/A	N/A	0	
i:4	CA	Ca	1	87.3 (45.1)	O_1	0.09	1	Poorly Coordinated	N/A	N/A	0	
i:6	CA	Ca	1	78.2 (62.6)	O_2	0.4	0.81	Square Planar	8.6°	50%	0	
i:8	CA	Ca	1	68.5 (38.8)	O_4	1	0.36	Trigonal Bipyramidal	9.8°	20%	0	
i:9	CA	Ca	1	81.6 (53.6)	O_2	0.4	0.84	Square Planar	23°	50%	0	
i:10	CA	Ca	1	77.3 (40.6)	O_1	0.19	1	Poorly Coordinated	N/A	N/A	0	
i:12	CA	Ca	1	57.5 (34.6)	O_2	0.3	0.71	Square Planar	0.3°	50%	0	
i:13	CA	Ca	1	84.4 (38.8)	O_1	0.09	1	Poorly Coordinated	N/A	N/A	0	
i:16	CA	Ca	1	42.4 (37.3)	O_6	1.3	0.13	Octahedral	19.2°	0	0	
i:17	CA	Ca	1	38.7 (39.4)	O_7	1.6	0.043	Pentagonal Bipyramidal	7.1°	0	0	
i:18	CA	Ca	1	37.4 (38.1)	O_7	1.3	0.13	Pentagonal Bipyramidal	8.3°	0	0	
i:19	CA	Ca	1	39.3 (37.8)	O_6	1.5	0.14	Octahedral	19.3°	0	0	Na
i:20	CA	Ca	1	42.3 (40.5)	O_7	1.4	0.045	Pentagonal Bipyramidal	7°	0	0	
i:21	CA	Ca	1	46 (43.4)	O_6	1.2	0.17	Octahedral	16.3°	0	0	
i:22	CA	Ca	1	41.2 (42.8)	O_7	1.5	0.099	Pentagonal Bipyramidal	6.5°	0	0	
i:23	CA	Ca	1	40.4 (39.9)	O_6	1.2	0.16	Octahedral	19.7°	0	0	
i:24	CA	Ca	1	45.3 (35.7)	O_3	0.6	0.69	Octahedral	23.4°	50%	0	K
i:25	CA	Ca	1	44.7 (40.4)	O_7	1.7	0.16	Pentagonal Bipyramidal	7.2°	0	0	
i:26	CA	Ca	1	45.6 (40.8)	O_5	1.1	0.28	Octahedral	10.2°	16%	0	
i:27	CA	Ca	1	41.4 (37.3)	O_6	1.5	0.081	Octahedral	18.3°	0	0	Na
i:28	CA	Ca	1	46.9 (46.9)	O_7	1.6	0.1	Pentagonal Bipyramidal	7.1°	0	0	
i:29	CA	Ca	1	46.7 (43)	O_6	1.3	0.33	Octahedral	15.6°	0	0	
i:30	CA	Ca	1	48.5 (46.2)	O_4	1.2	0.41	Octahedral	6.7°	33%	0	
i:31	CA	Ca	1	54.5 (47.7)	O_5	1.1	0.26	Trigonal Bipyramidal	14.4°	0	0	
i:33	CA	Ca	1	76.2 (47)	O_3	0.9	0.4	Square Planar	14.7°	25%	0	
i:34	CA	Ca	1	88.9 (57.1)	O_2	0.3	0.83	Square Planar	15.1°	50%	0	
Legend:			Not applicable	Outlier	Borderline	Acceptable						

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distributions for 0973sH_q_S200_refine_64.pdb in comparison with CSD



(1) Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) *Nature Protocols, 9(1)*, 156-70.
(2) Brown ID (2009) *Chem. Rev., 109*, 6858-6919.
(3) Müller P, Köpke S, Sheldrick GM (2003) *Acta Crystallogr. D Biol. Crystallogr., 59*, 32-37.
(4) Kuppuraj G, Dudev M, Lim C (2009) *J. Phys. Chem. B, 113*, 2952-2960.
(5) *CSD: Cambridge Structural Database*
Maintained by: Heping Zheng <dust@iwonka.med.virginia.edu>

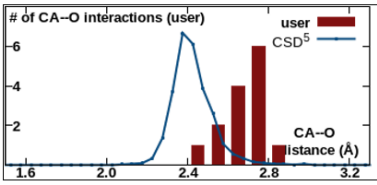
Citing CheckMyMetal (CMM):
Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. Zheng,H., Chordia,M.D., Cooper,D.R., Chruszcz,M., Müller,P., Sheldrick,G.M., Minor,W. (2014) *Nature Protocols, 9(1)*, 156-70.

ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
f:2	CA	Ca	1	66.6 (57.3)	O_1	0.3	1	Poorly Coordinated	N/A	N/A	0	
f:5	CA	Ca	1	78.2 (71.5)	O_2	0.3	0.65	Tetrahedral	7.5°	50%	0	
f:9	CA	Ca	1	*85.5 (65.4)*	O_2	0.3	0.84	Square Planar	23.3°	50%	0	
f:10	CA	Ca	1	81.8 (87.3)	O_1	0.15	1	Poorly Coordinated	N/A	N/A	0	
f:11	CA	Ca	1	*80.6 (53.4)*	O_1	0.15	1	Poorly Coordinated	N/A	N/A	0	
f:12	CA	Ca	1	*80.1 (66.9)*		N/A	N/A	Free	N/A	N/A	N/A	
f:13	CA	Ca	1	*86.8 (72.2)*	O_2	0.3	0.72	Square Planar	2.1°	50%	0	
f:14	CA	Ca	*0.71*	*60.9 (46.4)*	O_2N_1	0.4	0.5	Tetrahedral	26.3°	*25%*	0	
f:16	CA	Ca	1	*80.4 (60.4)*	O_2	0.2	0.69	Square Planar	3.1°	50%	0	
f:17	CA	Ca	1	84.3 (85.5)	O_1	0.11	1	Poorly Coordinated	N/A	N/A	0	
f:18	CA	Ca	1	85.5 (77.7)		N/A	N/A	Free	N/A	N/A	N/A	
f:19	CA	Ca	1	*70.5 (84)*		N/A	N/A	Free	N/A	N/A	N/A	
f:20	CA	Ca	1	*85.7 (71.2)*		N/A	N/A	Free	N/A	N/A	N/A	

Legend: Not applicable | **Outlier** | *Borderline* | Acceptable

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distribution for 73sH_E32A_phenixserver_refine_10.pdb in comparison with CSD



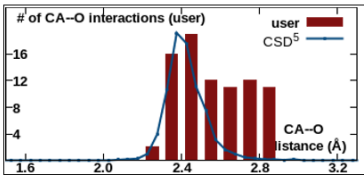
(1) Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) *Nature Protocols, 9(1)*, 156-70.
(2) Brown ID (2009) *Chem. Rev., 109*, 6858-6919.

ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
e:1	CA	Ca	1	60.8 (54.8)	O_5	1	0.29	Octahedral	20.3°	16%	0	
e:2	CA	Ca	1	56.4 (50.6)	O_5	1	0.25	Trigonal Bipyramidal	18°	0	0	
e:3	CA	Ca	1	56.8 (53.4)	O_3	0.7	0.15	Square Planar	15.2°	25%	0	K
e:4	CA	Ca	1	61.8 (53.8)	O_4	0.8	0.27	Square Planar	19.8°	0	0	
e:5	CA	Ca	1	60.2 (52.6)	O_6	1.3	0.083	Octahedral	20.6°	0	0	
e:6	CA	Ca	1	60.5 (53.7)	O_5	1	0.18	Octahedral	11°	16%	0	
e:7	CA	Ca	1	60.2 (57.6)	O_4	0.9	0.21	Square Planar	19.8°	0	0	
e:8	CA	Ca	1	57.4 (51)	O_5	1	0.32	Octahedral	14.7°	16%	0	
e:9	CA	Ca	1	61.3 (56.3)	O_5	1	0.31	Octahedral	21.2°	16%	0	
e:10	CA	Ca	1	59.3 (50.4)	O_4	0.7	0.3	Octahedral	12.6°	33%	0	K
e:11	CA	Ca	1	61.3 (55.9)	O_4	0.8	0.17	Square Planar	17.4°	0	0	
e:12	CA	Ca	1	59.8 (46)	O_3	0.5	0.61	Tetrahedral	41.5°	25%	0	
e:13	CA	Ca	1	60.8 (52.3)	O_5	0.8	0.23	Trigonal Bipyramidal	17.4°	0	0	
e:14	CA	Ca	1	58.6 (53.3)	O_5	1.1	0.34	Octahedral	14.8°	16%	0	
e:15	CA	Ca	1	64.8 (41.5)	O_2	0.6	0.77	Square Planar	8.1°	50%	0	K
e:17	CA	Ca	1	71.1 (64.2)	O_2	0.4	0.14	Linear	12.7°	0	0	
e:19	CA	Ca	0.9	65.7 (47.6)	O_5	1	0.57	Octahedral	16.5°	33%	1	
e:20	CA	Ca	0.72	56.7 (37.5)	O_3	1.1	0.4	Square Planar	9.2°	25%	0	
e:21	CA	Ca	1	69.1 (62.2)	O_2	0.4	0.89	Square Planar	22.1°	50%	0	
e:22	CA	Ca	1	70.3 (65.3)	O_3	0.6	0.13	Square Planar	13.6°	25%	0	K
e:23	CA	Ca	0.99	68.3 (46.7)	O_3	0.4	0.45	Square Planar	35.2°	25%	0	

Legend: Not applicable | **Outlier** | *Borderline* | Acceptable

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distribution for 73sH_E62A_Fesoak_Fe-phenixserver_refine_5.pdb in comparison with CSD



4. Rru-EncFtnSH-H65A

ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
g:1	CA	Ca	1	66.5 (59.7)		N/A	N/A	Free	N/A	N/A	N/A	
g:2	CA	Ca	1	84.7 (85.9)	O_2	0.2	0.45	Trigonal Planar	7.4°	33%	0	
g:3	CA	Ca	1	55.3 (69)	O_2	0.7	0.077	Linear	6.3°	0	0	
g:4	CA	Ca	1	58.1 (54.9)	O_3	1	0.36	Square Planar	22°	25%	0	
g:5	CA	Ca	1	69.1 (67.7)	O_3	0.7	0.6	Square Planar	35.2°	25%	0	K
g:6	CA	Ca	1	72.6 (57)	O_4	1	0.39	Tetrahedral	29.7°	0	0	
g:7	CA	Ca	1	66.8 (75.7)	O_2	0.2	0.52	Trigonal Planar	1.6°	33%	0	
g:8	CA	Ca	1	87.8 (68.5)	O_2	0.2	0.86	Square Planar	27°	50%	0	
g:9	CA	Ca	1	90.2 (66.5)		N/A	N/A	Free	N/A	N/A	N/A	
g:10	CA	Ca	1	60.3 (48.2)	O_3	0.7	0.39	Tetrahedral	21.4°	25%	0	K
g:11	CA	Ca	1	80.9 (60.5)		N/A	N/A	Free	N/A	N/A	N/A	
g:12	CA	Ca	1	88.6 (79.4)	O_2	0.5	0.38	Linear	35.4°	0	0	
g:13	CA	Ca	1	80.8 (79.4)	O_1	0.4	1	Poorly Coordinated	N/A	N/A	0	
g:14	CA	Ca	1	83.5 (59.1)	O_2	0.3	0.61	Trigonal Planar	13.5°	33%	0	
g:15	CA	Ca	1	70.4 (68.3)	O_2	0.4	0.19	Linear	17.6°	0	0	
g:16	CA	Ca	1	80.6 (73.6)	O_2	0.3	0.6	Trigonal Planar	3.3°	33%	0	
g:17	CA	Ca	1	83.6 (79)	O_1	0.13	1	Poorly Coordinated	N/A	N/A	0	
g:18	CA	Ca	1	84.9 (68.2)	O_1	0.1	1	Poorly Coordinated	N/A	N/A	0	
g:19	CA	Ca	1	53.5 (53.5)	O_3	0.5	0.36	Square Planar	29°	25%	0	K
g:20	CA	Ca	1	72 (53.9)	O_3	0.9	0.29	Tetrahedral	25°	25%	0	
g:21	CA	Ca	1	64.1 (55.6)	O_2	0.5	0.66	Trigonal Planar	5.9°	33%	0	
g:22	CA	Ca	1	64.4 (48.9)	O_3	0.5	0.26	Trigonal Planar	10°	0	0	
g:23	CA	Ca	1	62.9 (49.9)	O_3	0.7	0.37	Tetrahedral	29°	25%	0	K
g:24	CA	Ca	1	52.9 (48.4)	O_3	0.6	0.29	Square Planar	28.1°	25%	0	K
g:25	CA	Ca	1	78.4 (62.1)	O_2	0.4	0.53	Trigonal Planar	7.1°	33%	0	
g:26	CA	Ca	1	67.5 (62.9)	O_3	0.7	0.51	Tetrahedral	24.3°	25%	0	K
g:27	CA	Ca	1	64.7 (47.6)	O_3	0.5	0.37	Tetrahedral	18.7°	25%	0	
g:28	CA	Ca	1	67.3 (59.5)	O_3	0.7	0.64	Tetrahedral	28.4°	25%	0	K
g:29	CA	Ca	1	77 (51.5)	O_3	0.6	0.51	Tetrahedral	17.7°	25%	0	K
g:30	CA	Ca	1	73.7	$O_$	0.5	0.87	Square	29.1°	50%	0	

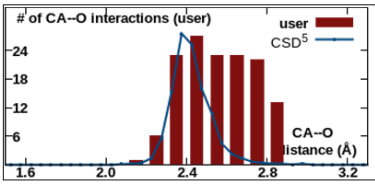
g:30	CA	Ca	1	*(61.7)*	O_2	**0.5**	**0.87**	**Planar**	**29.1°**	**50%**	0	
g:31	CA	Ca	1	*68.3 (58.2)*	O_2	**0.18**	*0.16*	**Linear**	*17.3°*	0	0	
g:32	CA	Ca	1	*74.8 (57.2)*	O_3	**0.7**	**0.42**	**Trigonal Planar**	*17.2°*	0	0	
g:33	CA	Ca	1	*70.6 (55.4)*	O_3	**0.7**	*0.21*	**Tetrahedral**	**24.5°**	*25%*	0	K
g:34	CA	Ca	1	*98.1 (81.8)*	O_2	**0.3**	**0.65**	**Tetrahedral**	7.1°	**50%**	0	
g:35	CA	Ca	1	*89.7 (76.9)*	O_2	**0.3**	**0.56**	**Trigonal Planar**	3.5°	**33%**	0	
g:36	CA	Ca	1	64 (73.6)	O_1	**0.17**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:37	CA	Ca	1	*55.2 (78.5)*	O_1	**0.16**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:38	CA	Ca	1	*61.3 (76.7)*	O_2	**0.5**	**0.28**	**Linear**	**26.9°**	0	0	
g:39	CA	Ca	1	63.7 (68.5)	O_2	**0.6**	*0.16*	**Linear**	13.2°	0	0	
g:40	CA	Ca	1	65.4 (68.1)	O_1	**0.1**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:41	CA	Ca	1	71.6 (74.4)	O_2	**0.4**	*0.18*	**Linear**	*17.1°*	0	0	
g:42	CA	Ca	1	62.2 (64.6)	O_2	**0.5**	**0.43**	**Linear**	11.8°	0	0	
g:43	CA	Ca	1	71.3 (73.4)	O_2	**0.4**	*0.17*	**Linear**	*17.9°*	0	0	
g:44	CA	Ca	1	65.1 (68.4)	O_2	**0.4**	**0.34**	**Linear**	**28.9°**	0	0	
g:45	CA	Ca	1	*66.2 (53.2)*	O_2	**0.2**	**0.62**	**Tetrahedral**	3.3°	**50%**	0	
g:46	CA	Ca	1	*67.1 (52.7)*	O_2	**0.7**	**0.75**	**Square Planar**	7.2°	**50%**	0	K
g:47	CA	Ca	1	74.3 (81.9)	O_2	**0.3**	**0.44**	**Linear**	**29.5°**	0	0	
g:48	CA	Ca	1	65.3 (67.3)	O_1	**0.14**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:49	CA	Ca	1	68.2 (67.9)	O_1	**0.19**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:50	CA	Ca	1	60.7 (58.4)	O_3	**0.5**	**0.56**	**Tetrahedral**	*15.2°*	*25%*	0	
g:51	CA	Ca	1	*72.9 (55.6)*	O_3	**0.8**	**0.51**	**Tetrahedral**	*13.9°*	*25%*	0	
g:52	CA	Ca	1	78.4 (69.5)	O_1	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:53	CA	Ca	1	74.9 (73.8)	O_1	**0.1**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:54	CA	Ca	1	69.6 (73)	O_1	**0.12**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:55	CA	Ca	1	73.7 (65)		N/A	N/A	**Free**	N/A	N/A	N/A	
g:56	CA	Ca	1	69.4 (69.8)	O_1	**0.13**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:57	CA	Ca	1	78.5 (68.6)	O_1	**0.14**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:58	CA	Ca	1	77.6 (68.5)	O_3	**0.4**	**0.37**	**Square Planar**	*20.3°*	*25%*	0	
g:59	CA	Ca	1	*78 (55.9)*		N/A	N/A	**Free**	N/A	N/A	N/A	
g:60	CA	Ca	1	*71.9 (59.5)*	O_1	**0.08**	**1**	**Poorly Coordinated**	N/A	N/A	0	
g:61	CA	Ca	1	72.9 (65.7)	O_2	**0.6**	**0.52**	**Trigonal Planar**	1.5°	**33%**	0	

g:62	CA	Ca	1	86.3 (68.1)	O_1	0.1	1	Poorly Coordinated	N/A	N/A	0	
g:63	CA	Ca	1	59.3 (42.4)	O_1	0.2	1	Poorly Coordinated	N/A	N/A	0	
g:64	CA	Ca	1	78.2 (56)	O_1	0.2	1	Poorly Coordinated	N/A	N/A	0	
g:65	CA	Ca	1	84.8 (61.4)	O_1	0.09	1	Poorly Coordinated	N/A	N/A	0	
g:66	CA	Ca	1	77.3 (66.1)	O_2	0.4	0.3	Linear	31.2°	0	0	
g:67	CA	Ca	1	62.1 (72.7)	O_2	0.7	0.16	Linear	11.6°	0	0	
g:68	CA	Ca	1	74.3 (56.5)	O_1	0.2	1	Poorly Coordinated	N/A	N/A	0	
g:69	CA	Ca	1	66.3 (55)	O_1	0.13	1	Poorly Coordinated	N/A	N/A	0	
g:70	CA	Ca	1	72.1 (60.1)	O_3	0.6	0.56	Square Planar	32°	25%	0	K
g:71	CA	Ca	1	79.7 (59.3)	O_2	0.4	0.41	Linear	29.5°	0	0	
g:72	CA	Ca	1	76.4 (49.1)	O_2	0.4	0.92	Square Planar	39.6°	50%	0	
g:73	CA	Ca	1	82.2 (47.8)	O_2	0.5	0.67	Square Planar	7.9°	50%	0	
g:74	CA	Ca	1	89.9 (49.8)	O_2N_1	0.9	0.48	Trigonal Bipyramidal	6°	40%	0	

Legend: Not applicable | **Outlier** | *Borderline* | Acceptable

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distribution for 0973sH_H65A_refine_21.pdb in comparison with CSD



(1) Zheng H, Chordia MD, Cooper DR, Chruszcz M, Müller P, Sheldrick GM, Minor W (2014) *Nature Protocols, 9(1)*, 156-70.
(2) Brown ID (2009) *Chem. Rev., 109*, 6858-6919.
(3) Müller P, Köpke S, Sheldrick GM (2003) *Acta Crystallogr. D Biol. Crystallogr., 59*, 32-37.
(4) Kuppuraj G, Dudev M, Lim C (2009) *J. Phys. Chem. B, 113*, 2952-2960.
(5) *CSD: Cambridge Structural Database*
Maintained by: Heping Zheng <dust@iwonka.med.virginia.edu>

5. Rru-EncFtnSH-E62D

Warning: *Valence* and *nVECSUM* parameters should be interpreted with great care due to the presence of multi-nuclear metal clusters around e:1 e:2 e:3 e:4 e:5 e:6 e:7 e:8 e:9 e:10 e:11 e:12 e:13 e:14 e:15 e:16 e:17 e:18 e:20 e:21 e:40 e:41 e:42 f:44 f:46 f:47 f:48

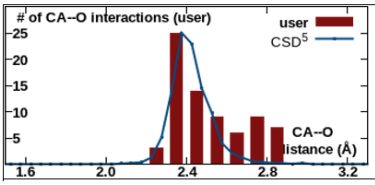
ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
e:1	FE	Fe	1	28.8 (27.8)	O_6	**0.9**	*0.16*	*Trigonal Bipyramidal*	*15°*	0	1	Na
e:2	FE	Fe	1	31.2 (28.1)	O_6	**0.9**	*0.18*	*Trigonal Bipyramidal*	*14.6°*	0	1	Na
e:3	FE	Fe	1	32.4 (28.8)	O_6N_1	**0.9**	0.095	*Octahedral*	13.4°	*16%*	2	
e:4	FE	Fe	1	*36.6 (29.8)*	O_6	**0.9**	*0.14*	*Trigonal Bipyramidal*	*14.4°*	0	1	Na
e:5	FE	Fe	1	30.4 (27.4)	O_6	**0.9**	*0.15*	*Trigonal Bipyramidal*	*14.6°*	0	1	Na
e:6	FE	Fe	1	29.6 (29.3)	O_6N_1	**0.9**	0.032	*Octahedral*	*14.3°*	*16%*	2	
e:7	FE	Fe	*0.98*	30.2 (29.5)	O_6N_1	**1**	*0.13*	*Octahedral*	*15.3°*	*16%*	2	
e:8	FE	Fe	*0.97*	31.3 (30)	O_6N_1	**1.1**	*0.13*	*Octahedral*	**21.9°**	0	1	
e:9	FE	Fe	1	29.8 (30.5)	O_6	**1**	*0.16*	*Trigonal Bipyramidal*	*13.8°*	0	1	Na
e:10	FE	Fe	*0.93*	27.2 (30.6)	O_6N_1	**1**	*0.13*	*Octahedral*	*16.2°*	*16%*	2	
e:11	FE	Fe	1	34.4 (36.2)	O_6	**0.9**	*0.12*	*Trigonal Bipyramidal*	*14.4°*	0	1	Na
e:12	FE	Fe	*0.93*	36.1 (36.2)	O_6N_1	**1.1**	*0.16*	*Octahedral*	*15.5°*	*16%*	2	
e:13	FE	Fe	*0.98*	34.5 (32.6)	O_6N_1	**1**	0.077	*Octahedral*	*15.8°*	*16%*	2	
e:14	FE	Fe	1	37.8 (32.7)	O_6N_1	**1**	0.08	*Octahedral*	*16.2°*	*16%*	2	
e:15	FE	Fe	1	34.3 (30.3)	O_6N_1	**1**	0.097	*Octahedral*	*14.1°*	*16%*	2	
e:16	FE	Fe	1	34.5 (33.4)	O_6N_1	**1.1**	*0.13*	*Octahedral*	*15.7°*	*16%*	2	
e:17	FE	Fe	1	36.4 (37.6)	O_6N_1	**1.1**	*0.13*	*Octahedral*	*16.3°*	*16%*	2	
e:18	FE	Fe	*0.88*	30.1 (34.8)	O_6N_1	**1.1**	*0.13*	*Octahedral*	*16.8°*	*16%*	2	
e:19	FE	Fe	1	38.7 (36.4)	O_6N_1	*1.2*	*0.14*	*Octahedral*	*16.9°*	*16%*	2	
e:20	FE	Fe	*0.87*	*39.8 (47.3)*	O_6	**1**	*0.18*	*Trigonal Bipyramidal*	*15.6°*	0	1	Na
e:21	FE	Fe	1	55.4 (52.4)	O_5	**0.9**	**0.3**	*Octahedral*	11.8°	**33%**	1	Na
e:40	FE	Fe	*0.97*	50 (50.1)	O_6	**1**	*0.15*	*Trigonal Bipyramidal*	13.5°	0	1	Na
e:41	FE	Fe	1	48.6 (48.5)	O_6N_1	**1.1**	*0.12*	*Octahedral*	*16°*	*16%*	2	
e:42	FE	Fe	*0.87*	31.4 (34.5)	O_6N_1	**1.1**	0.1	*Octahedral*	*16.9°*	*16%*	2	
e:43	FE	Fe	*0.91*	30.3 (34.6)	O_6N_1	*1.2*	*0.11*	*Trigonal Bipyramidal*	13°	0	2	
f:44	FE	Fe	1	33.2 (31.2)	O_6N_1	**1.1**	0.091	*Octahedral*	*16.5°*	*16%*	2	
f:45	FE	Fe	1	34.4 (33.1)	O_6N_1	*1.2*	*0.16*	*Octahedral*	*13.8°*	*16%*	2	
f:46	FE	Fe	1	39.7 (35.1)	O_6N_1	**1.1**	*0.12*	*Octahedral*	**21.8°**	0	1	

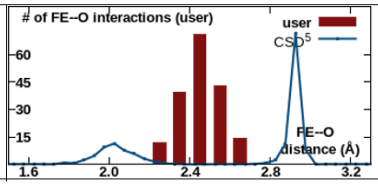
f:47	FE	Fe	0.99	31.9 (32.4)	O_6N_1	**1**	0.067	*Octahedral*	*16.5°*	*16%*	2	
f:48	FE	Fe	1	37.9 (36.9)	O_6	**1**	*0.2*	*Trigonal Bipyramidal*	*15.3°*	0	1	Na
j:1	CA	Ca	1	*70.1 (48.3)*	**O_1N_1**	**0.4**	**0.86**	**Square Planar**	*16.1°*	**50%**	0	
j:2	CA	Ca	1	*70.9 (44.8)*	**O_1N_1**	**0.3**	**0.62**	**Tetrahedral**	0.9°	**50%**	0	
j:3	CA	Ca	1	*73.9 (48.6)*		N/A	N/A	**Free**	N/A	N/A	N/A	
j:4	CA	Ca	1	*70 (44.1)*	*O_2*	**0.6**	**0.69**	**Square Planar**	3°	**50%**	0	K
j:5	CA	Ca	1	79.1 (72.9)	**O_1**	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:6	CA	Ca	1	*71.2 (55.3)*	**O_2N_1**	**0.4**	**0.61**	**Square Planar**	**26.6°**	*25%*	0	
j:7	CA	Ca	1	79.3 (69.1)	**O_1**	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:8	CA	Ca	1	*79.2 (60.6)*	**O_1N_1**	**0.3**	**0.86**	**Square Planar**	**27.6°**	**50%**	0	
j:9	CA	Ca	1	*74.8 (57.8)*	**O_1**	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:10	CA	Ca	1	*64.9 (53.7)*	**O_1N_1**	**0.3**	**0.82**	**Square Planar**	*15.8°*	**50%**	0	
j:11	CA	Ca	1	53 (50.8)	O_5	*1.4*	**0.25**	**Trigonal Bipyramidal**	*14.4°*	0	0	Na
j:12	CA	Ca	1	*68.2 (47.8)*	**O_1N_1**	**0.4**	**0.83**	**Square Planar**	*19.8°*	**50%**	0	
j:13	CA	Ca	1	*69.8 (49.2)*	**N_1**	**0.15**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:14	CA	Ca	1	64.9 (57.9)	**N_1**	**0.09**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:15	CA	Ca	1	*72.2 (56.2)*	**O_1**	**0.2**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:16	CA	Ca	1	*69.3 (45.2)*	**O_1N_1**	**0.2**	**0.84**	**Square Planar**	**21.8°**	**50%**	0	
j:17	CA	Ca	1	*71.4 (46.4)*	**O_1N_1**	**0.4**	**0.8**	**Square Planar**	*15.4°*	**50%**	0	
j:18	CA	Ca	1	75.7 (67.8)	**N_1**	**0.1**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:19	CA	Ca	1	*75.6 (54.1)*	**O_1N_1**	**0.2**	**0.83**	**Square Planar**	*20.5°*	**50%**	0	
j:20	CA	Ca	1	*68.7 (51.3)*	**O_1**	**0.2**	**1**	**Poorly Coordinated**	N/A	N/A	0	
j:21	CA	Ca	1	54.5 (56.9)	O_3	**0.8**	*0.21*	**Square Planar**	7.5°	*25%*	0	
j:22	CA	Ca	1	48.1 (50.3)	O_3	**0.8**	*0.19*	**Square Planar**	8.4°	*25%*	0	
j:23	CA	Ca	1	46.8 (51.8)	O_3	**0.8**	0.098	**Square Planar**	4.2°	*25%*	0	
j:24	CA	Ca	1	56.5 (59.1)	O_3	**0.7**	*0.19*	**Square Planar**	7.9°	*25%*	0	K
j:25	CA	Ca	1	*55.8 (47.6)*	O_4	**1**	**0.38**	Octahedral	10.6°	**33%**	0	
j:26	CA	Ca	1	*52 (64.2)*	*O_2*	**0.5**	*0.2*	**Linear**	*15.1°*	0	0	
j:27	CA	Ca	1	57.5 (59.3)	O_3	**0.8**	**0.27**	**Square Planar**	10.9°	*25%*	0	
j:28	CA	Ca	1	52 (51.5)	O_4	**0.9**	**0.34**	Octahedral	4°	**33%**	0	
j:29	CA	Ca	1	55.8 (60)	O_3	**0.7**	*0.13*	**Square Planar**	6.8°	*25%*	0	K

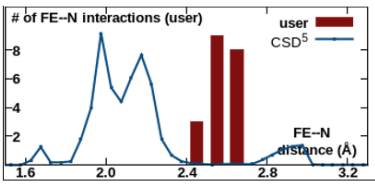
j:30	CA	Ca	1	60.2 (58.5)	O_2	0.5	0.19	Linear	21.4°	0	0	
j:31	CA	Ca	1	57.5 (56.3)	O_2	0.6	0.11	Linear	5.7°	0	0	
j:32	CA	Ca	1	55.1 (59.8)	O_3	0.9	0.14	Square Planar	9.4°	25%	0	
j:33	CA	Ca	1	70.5 (61.1)	O_2	0.6	0.13	Linear	15.2°	0	0	
j:34	CA	Ca	1	72.8 (55.4)	N_1	0.09	1	Poorly Coordinated	N/A	N/A	0	
j:35	CA	Ca	1	74.7 (48.8)	O_1	0.4	1	Poorly Coordinated	N/A	N/A	0	
j:36	CA	Ca	1	70.7 (54.8)	O_1N_1	0.4	0.78	Square Planar	13°	50%	0	
j:37	CA	Ca	1	72.4 (53.9)	O_1N_1	0.3	0.83	Square Planar	20.5°	50%	0	
j:38	CA	Ca	1	79.1 (70.8)		N/A	N/A	Free	N/A	N/A	N/A	
j:39	CA	Ca	1	69.5 (45.8)	O_2N_1	0.5	0.58	Square Planar	28.9°	25%	0	
j:40	CA	Ca	1	66.8 (45)	O_2	0.3	0.93	Square Planar	42.8°	50%	0	
k:9	CA	Ca	1	64.3 (59)	O_3	0.9	0.42	Square Planar	11.1°	25%	0	
l:10	CA	Ca	1	65.9 (56.6)	O_1N_1	0.2	0.76	Square Planar	7.5°	50%	0	
l:11	CA	Ca	1	76.2 (53.2)	O_1N_1	0.3	0.41	Trigonal Planar	12.8°	33%	0	
l:12	CA	Ca	1	68.9 (46.8)	O_2	0.6	0.81	Square Planar	17.4°	50%	0	K
	Legend:		Not applicable	Outlier	Borderline	Acceptable						

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distributions for 73sH_E62D_phenixserver_refine_14.pdb in comparison with CSD







6. Rru-EncFtnSH-E31A

Warning: *Valence* and *nVECSUM* parameters should be interpreted with great care due to the presence of multi-nuclear metal clusters around e:5 e:25 e:26 e:28

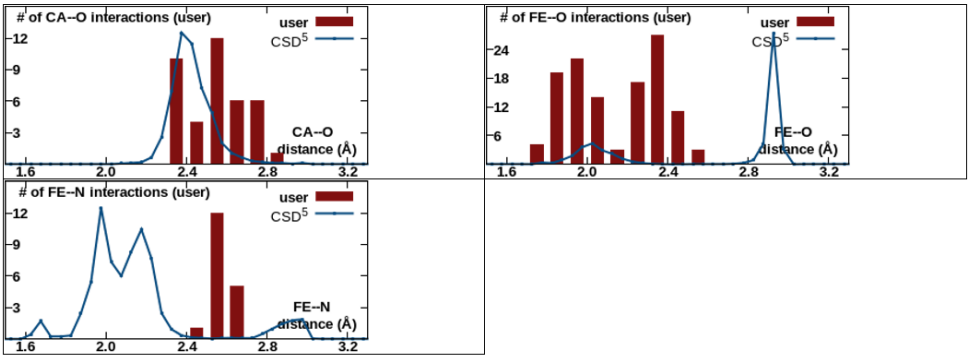
ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD*(°)[1]	*Vacancy*[1]	Bidentate	Alt. metal
e:1	FE	Fe	1	43.2 (41.2)	O_4N_1	1.7	0.13	Trigonal Bipyramidal	8.2°	20%	1	
e:2	FE	Fe	1	39.3 (34.8)	O_4	1.5	0.2	**Trigonal Planar**	11.8°	0	1	
e:3	FE	Fe	1	43 (41.5)	O_4N_1	1.8	0.16	Tetrahedral	19.1°	0	1	Cu
e:4	FE	Fe	1	39.6 (39.3)	O_4N_1	1.8	0.18	Trigonal Bipyramidal	10°	20%	1	
e:5	FE	Fe	1	44.3 (41.5)	O_4N_1	1.8	**0.24**	Trigonal Bipyramidal	8.8°	20%	1	
e:6	FE	Fe	1	42.8 (39.9)	O_4N_1	1.7	0.2	Tetrahedral	19.2°	0	1	
e:7	FE	Fe	0.87	32.7 (39.4)	O_4N_1	1.8	0.098	Tetrahedral	20.7°	0	1	Cu
e:8	FE	Fe	0.98	40.2 (40.7)	O_4N_1	1.8	0.16	Tetrahedral	16.4°	0	1	Cu
e:9	FE	Fe	1	44.6 (38.8)	O_4N_1	1.7	0.13	Trigonal Bipyramidal	8.6°	20%	1	
e:10	FE	Fe	0.79	36 (43.3)	O_4N_1	1.8	0.14	Trigonal Bipyramidal	7.5°	20%	1	
e:11	FE	Fe	0.94	41.4 (39.7)	O_4	1.7	0.14	**Trigonal Planar**	13.9°	0	1	
e:12	FE	Fe	0.84	39.1 (44.8)	O_4N_1	1.7	0.072	Trigonal Bipyramidal	8.7°	20%	1	
e:13	FE	Fe	0.88	55.3 (53)	O_4	1.8	0.18	**Trigonal Planar**	14.5°	0	1	Cu
e:14	FE	Fe	1	52.8 (53.5)	O_4	1.9	0.19	**Trigonal Planar**	11.5°	0	1	Cu
e:15	FE	Fe	1	39.2 (36.9)	O_4	1.9	0.15	**Trigonal Planar**	13.8°	0	1	Cu
e:16	FE	Fe	0.89	42.5 (40.7)	O_4	1.7	0.12	**Trigonal Planar**	11.9°	0	1	
e:17	FE	Fe	1	41.1 (39.1)	O_4	1.7	0.17	**Trigonal Planar**	15.4°	0	1	
e:18	FE	Fe	0.97	42.4 (42.2)	O_4N_1	1.8	0.11	Tetrahedral	18.2°	0	1	Cu
e:19	FE	Fe	0.87	32.9 (37)	O_4N_1	1.9	0.16	Tetrahedral	15.6°	0	1	Cu
e:20	FE	Fe	0.93	36.9 (37.1)	O_4N_1	1.9	0.06	Trigonal Bipyramidal	9.4°	20%	1	
e:21	FE	Fe	1	42.5 (37.7)	O_4N_1	2	0.22	Trigonal Bipyramidal	10.2°	20%	1	Co
e:22	FE	Fe	0.93	41.8 (42.9)	O_4	1.7	0.081	**Trigonal Planar**	11°	0	1	
e:23	FE	Fe	0.92	40.5 (42)	O_4N_1	1.7	0.074	Tetrahedral	18.8°	0	1	
e:24	FE	Fe	0.83	32.3 (38.3)	O_4N_1	1.8	0.16	Tetrahedral	20.5°	0	1	Cu
e:25	FE	Fe	0.87	40.7 (45.4)	O_4	1.7	**0.28**	**Trigonal Planar**	15.4°	0	1	
e:26	FE	Fe	0.89	44.5 (38.9)	O_4N_1	1.9	**0.25**	Tetrahedral	19.4°	0	1	Cu
e:27	FE	Fe	0.95	42.9 (49)	O_4	1.8	0.082	**Trigonal Planar**	7.3°	0	1	
e:28	FE	Fe	1	48.5 (49.9)	O_4	1.9	**0.33**	Tetrahedral	19°	25%	1	Cu
e:29	FE	Fe	1	62.1	O_4	1.7	0.16	**Trigonal**	17.4°	0	1	

e:29	FE	Fe	1	(57.8)	O_4	1.7	0.16	Planar	17.4	0	1	
e:30	FE	Fe	1	64.9 (56.2)	O_4N_1	1.9	0.15	Tetrahedral	17.3°	0	1	Cu
g:2	CA	Ca	1	69.3 (56.8)	O_1	0.17	1	Poorly Coordinated	N/A	N/A	0	
g:3	CA	Ca	1	73.3 (66.6)	O_2	0.4	0.66	Trigonal Planar	0.7°	33%	0	
g:4	CA	Ca	1	72.1 (50.1)	O_2	0.4	0.86	Square Planar	28.3°	50%	0	
g:5	CA	Ca	1	78.9 (64.7)	O_2	0.6	0.77	Square Planar	8.1°	50%	0	K
g:6	CA	Ca	1	81.3 (62.1)	N_1	0.14	1	Poorly Coordinated	N/A	N/A	0	
g:7	CA	Ca	1	78.3 (56.9)	O_3	0.6	0.6	Trigonal Bipyramidal	12°	40%	0	K
g:8	CA	Ca	1	65.1 (40.9)	O_3	0.8	0.74	Octahedral	23.1°	50%	0	
g:9	CA	Ca	1	70.3 (50.9)	O_2	0.4	0.53	Trigonal Planar	1.8°	33%	0	
g:10	CA	Ca	1	70.4 (52.9)	O_2	0.3	0.49	Trigonal Planar	2°	33%	0	
g:11	CA	Ca	1	73.5 (56.8)	O_1	0.13	1	Poorly Coordinated	N/A	N/A	0	
g:12	CA	Ca	1	76.5 (61.4)	N_1	0.2	1	Poorly Coordinated	N/A	N/A	0	
g:15	CA	Ca	1	79.5 (66.8)	O_2	0.3	0.57	Tetrahedral	5.1°	50%	0	
g:16	CA	Ca	1	77.1 (51.9)	O_3N_1	0.7	0.37	Square Planar	29.1°	0	0	
g:17	CA	Ca	1	73.4 (56.6)	O_1N_1	0.3	0.79	Square Planar	4.8°	50%	0	
g:18	CA	Ca	1	82.8 (55.8)	O_4	1	0.45	Trigonal Bipyramidal	8.5°	20%	0	
g:19	CA	Ca	1	85.5 (62)	O_3	0.6	0.57	Tetrahedral	26.3°	25%	0	K
g:20	CA	Ca	1	84.1 (67.3)	O_1	0.18	1	Poorly Coordinated	N/A	N/A	0	
g:21	CA	Ca	1	80.7 (56.8)	O_1	0.18	1	Poorly Coordinated	N/A	N/A	0	
g:22	CA	Ca	1	75 (56.6)	O_2	0.6	0.71	Square Planar	0.7°	50%	0	K
g:24	CA	Ca	1	85.3 (67.6)	O_2	0.5	0.63	Trigonal Planar	2.9°	33%	0	
g:25	CA	Ca	1	84 (69.9)	O_1	0.1	1	Poorly Coordinated	N/A	N/A	0	
g:26	CA	Ca	1	69.9 (39.2)	O_1	0.3	1	Poorly Coordinated	N/A	N/A	0	

Legend:	Not applicable	Outlier	*Borderline*	Acceptable

Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distributions for 73sH_E31A_phenixserver_refine_5.pdb in comparison with CSD



7. Rru-EncFtnSH-E34A

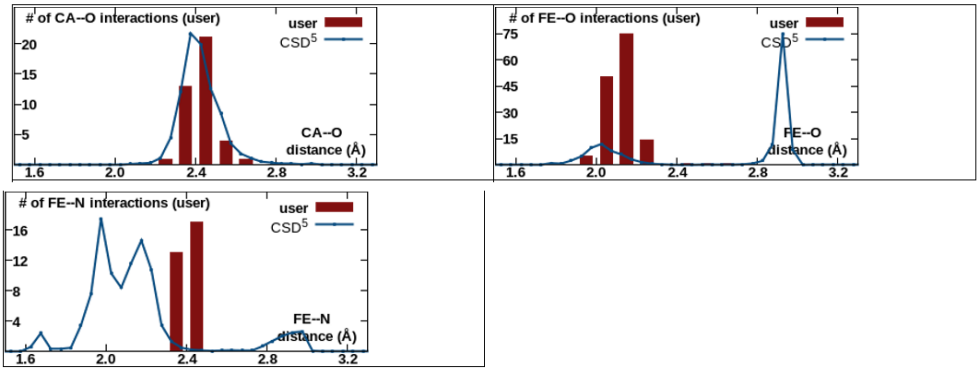
ID	Res.	Metal	Vaal encaE	C fnatM upcv·)[1]	t σ ncas	. nlecae[2]	c. EC SUN[3]	GeMnetry[1,4]	r RN gD(°)[1]	. ncncaE[1]	Bidentate	Alt. metal
e:1	FE	Fe	1	44.4 (32.2)	O_5N_1	2	0.13	Trigonal Bipyramidal	7.6°	0	1	Co
e:2	FE	Fe	0.77	32.9 (33.6)	O_5N_1	2.1	0.15	Trigonal Bipyramidal	5.7°	0	1	Co
e:3	FE	Fe	1	41.3 (31.6)	O_5N_1	1.9	0.17	Trigonal Bipyramidal	5°	0	1	Co
e:4	FE	Fe	1	41.1 (33.1)	O_5N_1	2.1	0.14	Trigonal Bipyramidal	6.6°	0	1	Co
e:5	FE	Fe	0.75	32.4 (32.2)	O_6N_1	2.2	0.14	Trigonal Bipyramidal	7.7°	0	2	Co
e:6	FE	Fe	1	41.3 (29.9)	O_5N_1	1.9	0.15	Trigonal Bipyramidal	7.5°	0	1	Co
e:7	FE	Fe	1	43.6 (34.6)	O_5N_1	2	0.14	Trigonal Bipyramidal	7.8°	0	1	Co
e:8	FE	Fe	1	42.3 (33.1)	O_5N_1	2	0.15	Trigonal Bipyramidal	5.8°	0	1	Co
e:9	FE	Fe	1	47.3 (37)	O_5N_1	1.9	0.15	Trigonal Bipyramidal	5.4°	0	1	Co
e:10	FE	Fe	1	43.2 (32.5)	O_5N_1	1.9	0.16	Trigonal Bipyramidal	7.3°	0	1	Co
e:11	FE	Fe	1	38.8 (32.2)	O_5N_1	1.9	0.15	Trigonal Bipyramidal	6.5°	0	1	
e:12	FE	Fe	0.67	32.4 (33.7)	O_5N_1	2.1	0.19	Trigonal Bipyramidal	9°	0	1	Co
e:13	FE	Fe	0.74	34.3 (31.6)	O_5N_1	2.1	0.14	Trigonal Bipyramidal	8.2°	0	1	Co
e:14	FE	Fe	1	52 (34.9)	O_5N_1	2	0.12	Trigonal Bipyramidal	4.3°	0	1	Co
e:15	FE	Fe	1	54 (38.1)	O_5N_1	2	0.14	Trigonal Bipyramidal	7.4°	0	1	Co
e:16	FE	Fe	1	49.4 (35.8)	O_5N_1	2.1	0.14	Trigonal Bipyramidal	7.2°	0	1	Co
e:17	FE	Fe	1	46.3 (34.7)	O_5N_1	2.1	0.12	Trigonal Bipyramidal	6.4°	0	1	Co
e:18	FE	Fe	1	45.8 (36.6)	O_5N_1	2	0.13	Octahedral	**21.6°**	0	0	Co, Cu
e:19	FE	Fe	1	43.3 (35.8)	O_5N_1	2.1	0.15	Trigonal Bipyramidal	9.1°	0	1	Co
e:20	FE	Fe	1	53.4 (41.5)	O_5N_1	2.2	0.17	Trigonal Bipyramidal	6.6°	0	1	Co
e:21	FE	Fe	1	49.3 (37.8)	O_5N_1	2.1	0.17	Trigonal Bipyramidal	6.5°	0	1	Co
e:22	FE	Fe	0.74	54 (43.1)	O_4N_1	2	0.065	Trigonal Bipyramidal	6.7°	20%	1	Co
e:23	FE	Fe	1	56.1 (42.8)	O_5N_1	2.2	0.15	Trigonal Bipyramidal	6.6°	0	1	Co
e:24	FE	Fe	1	58 (42.4)	O_4N_1	1.7	0.11	Trigonal Bipyramidal	6.8°	20%	1	
e:25	FE	Fe	0.75	54.2 (47.2)	O_4N_1	2	0.079	Tetrahedral	12.5°	0	1	
e:26	FE	Fe	1	44.8 (34.6)	O_5N_1	2.1	0.15	Trigonal Bipyramidal	6.2°	0	1	Co
e:27	FE	Fe	1	67.6 (49.3)	O_4N_1	1.8	0.068	Trigonal Bipyramidal	7.3°	20%	1	
e:28	FE	Fe	1	42 (32.9)	O_6N_1	1.9	0.064	Octahedral	**26.4°**	0	1	Cu
e:29	FE	Fe	1	57.5 (45.3)	O_4N_1	1.9	0.058	Trigonal Bipyramidal	7°	20%	1	Co
e:30	FE	Fe	1	51.1	O_5N_1	2.1	0.13	Trigonal	7.3°	0	1	Co

e:30	FE	Fe	1	*(41.3)*	O_5N_1	2.1	*0.13*	*Bipyramidal*	7.3	0	1	Co
h:4	CA	Ca	1	*54 (40.1)*	O_6	1.7	*0.22*	**Trigonal Bipyramidal**	13.5°	0	1	Na
h:5	CA	Ca	1	*72.6 (43.9)*	O_2	**0.6**	**0.81**	**Square Planar**	*19°*	**50%**	0	K
h:6	CA	Ca	1	*76.2 (62.9)*	O_1	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
h:7	CA	Ca	1	*72.4 (51)*	O_3	**1**	**0.35**	**Square Planar**	11.6°	*25%*	0	
h:10	CA	Ca	1	*52.5 (37.8)*	O_4	**0.9**	**0.39**	Octahedral	*18.5°*	**33%**	0	
h:11	CA	Ca	*0.87*	*80.6 (47.9)*	O_3	**0.8**	**0.33**	**Square Planar**	**36.7°**	*25%*	0	
h:12	CA	Ca	1	*77.2 (54.4)*	O_2	**0.6**	**0.77**	**Square Planar**	10.1°	**50%**	0	K
h:13	CA	Ca	1	*63.8 (39)*	O_4	**1**	**0.41**	**Tetrahedral**	11.7°	*25%*	1	
h:14	CA	Ca	1	*82.4 (50.4)*	O_3	**0.9**	**0.64**	Octahedral	10.5°	**50%**	0	
h:15	CA	Ca	1	*69.3 (43.9)*	O_2	**0.6**	**0.75**	**Square Planar**	6.4°	**50%**	0	K
h:16	CA	Ca	1	*80.2 (52.6)*	O_2	**0.5**	**0.81**	**Square Planar**	*18.4°*	**50%**	0	K
h:17	CA	Ca	1	*70.3 (45.6)*	O_2	**0.6**	**0.72**	**Square Planar**	2.6°	**50%**	0	K
h:18	CA	Ca	1	*74.5 (58.1)*	O_1	**0.3**	**1**	**Poorly Coordinated**	N/A	N/A	0	
h:19	CA	Ca	1	**87.1 (44.8)**	O_3N_1	*1.5*	**0.27**	**Square Planar**	*18.2°*	0	0	
h:20	CA	Ca	1	*86.7 (50)*	O_2	**0.5**	**0.85**	**Square Planar**	**27.3°**	**50%**	0	

Legend: Not applicable | **Outlier** | *Borderline* | Acceptable

Column	Description
Occupancy	Occupancy of ion under consideration
B factor (env.)[1]	Metal ion B factor, with valence-weighted environmental average B factor in parenthesis
Ligands	Elemental composition of the coordination sphere
Valence[2]	Summation of bond valence values for an ion binding site. *Valence* accounts for metal-ligand distances
nVECSUM[3]	Summation of ligand vectors, weighted by bond valence values and normalized by overall valence. Increase when the coordination sphere is not symmetrical due to incompleteness.
Geometry[1,4]	Arrangement of ligands around the ion, as defined by the *NEIGHBORHOOD* algorithm
gRMSD(°)[1]	R.M.S. Deviation of observed geometry angles (L-M-L angles) compared to ideal geometry, in degrees
Vacancy[1]	Percentage of unoccupied sites in the coordination sphere for the given geometry
Bidentate	Number of residues that form a bidentate interaction instead of being considered as multiple ligands
Alt. metal	A list of alternative metal(s) is proposed in descending order of confidency, assuming metal environment is accurately determined. This feature is still experimental. It requires user discrimination and cannot be blindly accepted

Metal-ligand distance distributions for 73sH_E34A_phenixserver_refine_6.pdb in comparison with CSD



8. Pfc-EncFtnS

ID	Res.	Metal	*Occupancy*	*B factor (env.)*[1]	*Ligands*	*Valence*[2]	*nVECSUM*[3]	*Geometry*[1,4]	*gRMSD(°)*[1]	*Vacancy*[1]	Bidentate	Alt. metal
e:1	FE	Fe	*0.8*	23 (23.3)	O_5N_1	2.3	0.078	*Trigonal Bipyramidal*	6.8°	0	1	Co, Zn
e:2	FE	Fe	*0.76*	21.7 (23.6)	O_5N_1	2.3	0.084	*Trigonal Bipyramidal*	7.6°	0	1	Co, Zn
e:3	FE	Fe	*0.63*	*18.8 (23.8)*	O_5N_1	2.2	0.083	*Trigonal Bipyramidal*	8.8°	0	1	Co
e:4	FE	Fe	*0.79*	22.8 (22.3)	O_5N_1	2.2	0.095	*Trigonal Bipyramidal*	5.9°	0	1	Co
e:5	FE	Fe	*0.7*	*21.1 (25)*	O_5N_1	2.1	0.086	*Trigonal Bipyramidal*	6.5°	0	1	Co
e:6	FE	Fe	*0.73*	23.5 (23.6)	O_5N_1	2.2	0.064	*Trigonal Bipyramidal*	9.1°	0	1	Co, Zn
e:7	FE	Fe	*0.79*	23.1 (22.9)	O_5N_1	2.2	*0.11*	*Trigonal Bipyramidal*	6.9°	0	1	Co, Zn
e:8	FE	Fe	*0.77*	21.1 (22.4)	O_5N_1	2.3	0.078	*Trigonal Bipyramidal*	6.2°	0	1	Co, Zn
e:9	FE	Fe	*0.74*	22.3 (24.7)	O_5N_1	2.2	*0.11*	*Trigonal Bipyramidal*	8.6°	0	1	Co, Zn
e:10	FE	Fe	*0.64*	*19.3 (22.7)*	O_5N_1	2.4	*0.13*	*Trigonal Bipyramidal*	8.3°	0	1	Co, Zn
e:11	FE	Fe	*0.71*	22.3 (23)	O_5N_1	2.1	0.094	*Trigonal Bipyramidal*	7.4°	0	1	Co
e:12	FE	Fe	*0.65*	20.5 (22.4)	O_5N_1	2.1	0.096	*Trigonal Bipyramidal*	6.4°	0	1	Co
e:13	FE	Fe	*0.63*	*22.1 (26.6)*	O_5N_1	2.3	0.1	*Trigonal Bipyramidal*	8.3°	0	1	Co, Zn
e:14	FE	Fe	*0.78*	24.1 (23.7)	O_5N_1	2.2	0.096	*Trigonal Bipyramidal*	7.2°	0	1	Co, Zn
e:15	FE	Fe	*0.8*	26.3 (25.8)	O_5N_1	2.2	*0.12*	*Trigonal Bipyramidal*	8.1°	0	1	Co, Zn
e:16	FE	Fe	*0.7*	18.1 (20.9)	O_5N_1	2.4	*0.11*	*Trigonal Bipyramidal*	8.3°	0	1	Co, Zn
e:17	FE	Fe	*0.64*	19.7 (22.6)	O_5N_1	2.3	*0.12*	*Trigonal Bipyramidal*	9.1°	0	1	Co, Zn
e:18	FE	Fe	*0.69*	22.4 (24.8)	O_5N_1	2.3	*0.13*	*Trigonal Bipyramidal*	7.7°	0	1	Co, Zn
e:19	FE	Fe	*0.66*	*16.7 (21.3)*	O_5N_1	2.3	0.084	*Trigonal Bipyramidal*	7.7°	0	1	Co, Zn
e:20	FE	Fe	*0.75*	25.5 (24.3)	O_5N_1	2.2	0.063	*Trigonal Bipyramidal*	7.5°	0	1	Co, Zn
e:21	FE	Fe	*0.62*	*19.8 (24.7)*	O_5N_1	2.3	0.052	*Trigonal Bipyramidal*	5.4°	0	1	Co, Zn
e:22	FE	Fe	*0.72*	26.1 (25.6)	O_5N_1	2.2	0.064	*Trigonal Bipyramidal*	8.8°	0	1	Co, Zn
e:23	FE	Fe	*0.7*	22.5 (25.2)	O_5N_1	2.3	0.081	*Trigonal Bipyramidal*	7.2°	0	1	Co, Zn
e:24	FE	Fe	*0.7*	26 (28.4)	O_5N_1	2.2	*0.11*	*Trigonal Bipyramidal*	7.1°	0	1	Co, Zn
e:25	FE	Fe	*0.72*	25 (27.2)	O_5N_1	2.1	0.097	*Trigonal Bipyramidal*	8.6°	0	1	Co
e:26	FE	Fe	*0.71*	24.4 (28)	O_5N_1	2.2	*0.13*	*Trigonal Bipyramidal*	7.9°	0	1	Co
e:27	FE	Fe	*0.62*	22.6 (25.1)	O_5N_1	2.4	*0.17*	*Trigonal Bipyramidal*	7.5°	0	1	Co, Zn
e:28	FE	Fe	*0.68*	28 (28.4)	O_5N_1	2.1	*0.11*	*Trigonal Bipyramidal*	8.9°	0	1	Co
e:29	FE	Fe	*0.82*	*32.1 (27.5)*	O_4N_1	2	0.09	Tetrahedral	10.8°	0	1	Cu, Co
e:30	FE	Fe	*0.62*	23 (26.4)	O_5N_1	2.2	0.068	*Trigonal Bipyramidal*	6.9°	0	1	Co, Zn

Legend: Not applicable | **Outlier** | *Borderline* | Acceptable